

NUMBER OF SEQUENCES: 36
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
 ADDRESSEE: P.O.
 STREET: 1755 Jefferson Davis Highway, Fourth Floor
 CITY: Arlington
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22202
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/263,413
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/882,980
 FILING DATE: 14-MAY-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Iswall, J. P.
 REGISTRATION NUMBER: 31,451
 REFERENCE/DOCKET NUMBER: 660-038-55X
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703)521-4500
 TELEFAX: (703)486-2347
 TELEX: 248855 OPAT UR
 INFORMATION FOR SEQ ID NO: 23:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 496 base pairs
 type: nucleic acid
 STRANDEDNESS: unknown
 TOPOLOGY: unknown
 MOLECULE TYPE: DNA (genomic)
 US-08-263-413-23

```

Query Match      2.8%; Score 42; DB 1; Length 496;
Best Local Similarity 65.2%; Pred. No. 0.0093;
Matches 60; Conservative 1; Mismatches 31; Indels 0; Gaps 0;

QY      1  GAYGGGAGCAGGTGTAACTGGGCGACAAAGATGACACCATGAGCACCACCGAT 60
          || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       250 GTCGATGAGATGAGAACGATGAGAGATGATGAGATGACTACTACGACGACGAGCAGC 309

QY      61  GATCACCATGACGACCATGATGATGATGATGA 92
          || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       310 GACGACGATGCTTCTTATGATGATGAGGATGA 341

RESULT 3
US-08-263-413-22
: Sequence 22, Application US/08263413
: Patent No. 5747246
: GENERAL INFORMATION:
: APPLICANT: PANNETIER, CHRISTOPHE
: APPLICANT: COCHET, MADELEINE
: APPLICANT: DARGHE, SYLVIE
: APPLICANT: KOURILSKY, PHILIPPE
: TITLE OF INVENTION: PROCESS FOR DETERMINING THE QUANTITY OF
: TITLE OF INVENTION: A DNA FRAGMENT OF INTEREST BY A METHOD OF
: TITLE OF INVENTION: METHOD OF ENZYMATIC AMPLIFICATION OF DNA
: NUMBER OF SEQUENCES: 36
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,
: ADDRESSEE: P.C.
: STREET: 1755 Jefferson Davis Highway, Fourth Floor
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202
: COMPUTER READABLE FORM:

```

```

? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/263,413
? FILING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/882,980
? FILING DATE: 14-MAY-1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Ivaldi & J.P.
? REGISTRATION NUMBER: 31,451
? REFERENCE/DOCKET NUMBER: 660-058-55X
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (703)521-4500
? TELEFAX: (703)466-2347
? TELEX: 248955 OFR TOR
? INFORMATION FOR SEQ ID NO: 22:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 500 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: unknown
? TOPOLOGY: unknown
? MOLECULE TYPE: DNA (genomic)
? US-08-263-413-22

```

```

Query Matchc 2.8%; Score 42; DB 1; Length 500;
Best Local Similarity 65.2%; Pred. No. 0.0093;
Matches 60; Conservative 1; Mismatches 31; Indels 0; Gaps 0;

QY      1  GAVGGGAGCAGCTGTCAGCATGGCGAAGCAAAAGATGACCCACCTGACGACGACCAACGAT 60
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      254  GTGATGATGAGATGAACACCATGAGCATGATGATGATGATGATGATGATGATGATGATG 313
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY      61  GATCACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 92
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      314  GACGACGATGCTTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 345

RESULT 4
US-07-807-043B-2
; Sequence 2, Application US/07807043B
; Patent No. 5342774
; GENERAL INFORMATION:
; APPLICANT: Boon, Thierry, Van den Eynde, Beno t
; TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
; TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felle & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/807,043B
; FILING DATE: 19911212
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702

```



```

1 REGISTRATION NUMBER: 30,946
2 REFERENCE/DOCKET NUMBER: LUD 5253.4-US
3 TELECOMMUNICATION INFORMATION:
4 TELEPHONE: (212) 688-9200
5 TELEFAX: (212) 838-3884
6 INFORMATION FOR SEQ ID NO: 2:
7 SEQUENCE CHARACTERISTICS:
8     LENGTH: 675 base pairs
9     TYPE: nucleic acid
10    STRANDEDNESS: single
11    TOPOLOGY: linear
12    MOLECULE TYPE: genomic DNA
13    OS-08-142-368A-2

```

```

Query Match: 2.8%; Score 42; DB 2; Length 675;
Best Local Similarity: 65.2%; Pred. No. 0.01;
Matches 60; Conservative 1; Mismatches 31; Indels 0; Gaps 0;

QY      1  GATGGGGAGAGATGTACGATGGGGGCGAAGAGTGGCCACATGACGACCGACGAT 60
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      244  GTCTGATGAGGATTAACACCATGACATCTGATGATGCTCTCTGACGACGAGGAC 303

QY      61  GATCACCATGACACCATGATGATGATGA 92
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      304  GACACGATCCCTTCTATGATGATGATGAGATGA 335

```

RESULT 7
 US-08-967-727-2
 : Sequence 2, Application US/08967727
 : Patent No. 6025474
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Gauglier, B atrice; Van den Eynde, Beno t;
 : TITLE OF INVENTION: van der Bruggen, Pierre; Boon-Failleux, Thierry for
 : TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding for
 : NUMBER OF SEQUENCES: 30
 : CORRESPONDENCE ADDRESSES:
 : ADDRESSEE: Felfe & Lynch
 : STREET: 805 Third Avenue
 : CITY: New York City
 : STATE: New York
 : ZIP: 10022
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
 :
 : COMPUTER: IBM
 : OPERATING SYSTEM: PC-DOS
 : SOFTWARE: Wordperfect
 :
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/967, 727
 : FILING DATE:
 : CLASSIFICATION: 435
 :
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 08/037, 230
 : FILING DATE: 26-MARCH-1993
 : APPLICATION NUMBER: PCT/US92/04354
 : FILING DATE: 22-MAY-1992
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 07/807, 0403
 : FILING DATE: 12-DECEMBER-1991
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 07/764, 365
 : FILING DATE: 23-SEPTEMBER-1991
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 07/728, 838
 : FILING DATE: 9-JULY-1991
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 07/705, 702
 : FILING DATE: 23-MAY-1991
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Hanson, No. 6025474man D.
 : REGISTRATION NUMBER: 30,946
 : REFERENCE/DOCKET NUMBER: LUD 5353

```
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 675 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
IS-O8-967-727-2
```

```

Query Match: 2.8%; Score 42; DB 3; Length 675;
Best Local Similarity: 65.2%; Pred. No. 0.011;
Matches 60; Conservative 1; Mismatches 31; Indels 0; Gaps 0.

QY 1 GATGGGAGAGATGTACGATGGGCGAAGAGTGCACCCATGACGACCAACGAT 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 244 GTCGATGAGGATTAACACGATGAGATGATGATGATGATGATGATGATGATGAT 303
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 GATCACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATGAT 92
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 304 GACGACGATGACCTTCTATGATGATGATGATGATGATGATGATGATGATGATGAT 335
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 8
 US-08-037-230D-2
 : Sequence 2, Application US/08037230D
 : Patent No. 6235525
 :
 : GENERAL INFORMATION:
 : APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;
 : APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry for
 : TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding for
 : TITLE OF INVENTION: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
 : NUMBER OF SEQUENCES: 30
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Felfe & Lynch
 : STREET: 805 Third Avenue
 : CITY: New York City
 : STATE: New York
 : ZIP: 10022
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
 :
 : COMPUTER: IBM
 : OPERATING SYSTEM: PC-DOS
 : SOFTWARE: Wordperfect
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/037, 230D
 : FILING DATE: 26-MARCH-1993
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: PCT/US92/04354
 : FILING DATE: 22-MAY-1992
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: PCT/807,043
 : FILING DATE: 12-DECEMBER-1991
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 07/764,364
 : FILING DATE: 23-SEPTEMBER-1991
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 07/728,838
 : FILING DATE: 9-JULY-1991
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 07/705,702
 : FILING DATE: 23-MAY-1991
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Hanson, No. 6235525man D.
 : REGISTRATION NUMBER: 30,946
 : REFERENCE/DOCKET NUMBER: LUD 5353
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (212) 688-9200
 : TELEFAX: (212) 838-3884
 :
 : INFORMATION FOR SEQ ID NO: 2:


```

SEQUENCE CHARACTERISTICS:
LENGTH: 675 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-037-230D-2

Query Match
Best Local Similarity 65.2%; Pred. No. 0.011;
Matches 60; Conservative 1; Mismatches 31; Indels 0; Gaps 0;

OY 1 GATGGGGCACTGTATGATGGCGAGAAACATCCATGATGCGACGACGAT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 244 GTCCGATGAGGATGAGAGACGATGAGATGATGAGATGATGATGATGAT 303
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 GATCACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGAT 92
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 304 GACGACGATGCTCTGATGATGATGATGATGATGATGATGATGATGATGAT 335
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
US-07-807-043B-4
Sequence 4, Application US/07807043B
Patent No. 5342774
GENERAL INFORMATION:
APPLICANT: Boon, Thierry, Van den Eynde, Beno t
TITLE OF INVENTION: Tumor Rejection Antigen Precursor, Tumor
TITLE OF INVENTION: Tumor Rejection Antigen Precursor, Tumor
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/807,043B
FILING DATE: 19911212
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,364
FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA: 07/728,838
FILING DATE: 09-JULY-1991
PRIOR APPLICATION DATA: 07/705,702
FILING DATE: 23-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5342774man D.
REGISTRATION NUMBER: 30,946
TELECOMMUNICATION INFORMATION: LUD 253.3
TELEPHONE: (212) 838-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1365 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: singular
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-07-807-043B-4

Query Match
Best Local Similarity 65.2%; Pred. No. 0.016;
Matches 60; Conservative 1; Mismatches 31; Indels 0; Gaps 0;

```

[illegible]

Mon Dec 2 12:19:30 2002

us-09-868-760-6.rni

Page 8

```

: NUMBER OF SEQUENCES: 48
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Felte & Lynch
: STREET: 805 Third Avenue
: CITY: New York City
: STATE: New York
: ZIP: 10022
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
: COMPUTER: IBM
: OPERATING SYSTEM: PC-DOS
: SOFTWARE: Mordperfect
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/299,849B
: FILING DATE: 1-SEPTEMBER-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/037,230
: FILING DATE: 26-MARCH-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US92/04354
: FILING DATE: 22-MAY-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/807,043
: FILING DATE: 12-DECEMBER-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/764,364
: FILING DATE: 23-SEPTEMBER-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/728,838
: APPLICATION NUMBER: 9-JULY-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/705,702
: FILING DATE: 23-MAY-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Hanson, No. 5612201man D.
: REGISTRATION NUMBER: 30,946
: REFERENCE/DOCKET NUMBER: LUD 5355
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 688-9200
: TELEFAX: (212) 838-3884
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4698 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: genomic DNA
: US-08-299-849B-5

Query Match 2.88; Score 42; DB 1; Length 4698;
Best Local Similarity 65.28; Pred. No. 0.031;
Matches 60; Conservative 1; Mismatches 31; Indels 0; Gaps 0;

OY 1 GAYGGGAGCGAGTCTAAGATGGGAGACCAAGATGACACCATGACGACCAACCATGAT 60
DB 706 GTGATGAGATGAGTGAAGAGATGATGATGATGATGATGATGATGATGATGATGAT 60
OY 61 GATCAGCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 92
DB 766 GACGACGATGCTCTCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 797
```

Search completed: November 29, 2002, 07:42:03
Job time : 95 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2002, 12:43:55 ; Search time 19 seconds
(without alignments)
769,642 Million cell updates/sec

Title: US-09-868-760-7

Perfect score: 497
Sequence: 1 DGEQNDGQNKDHHDDHD.....MOGHEVESERACVIGRA 497

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2.6/ptodata/1/aa/5A.COMB.pep.*
2: /cgn2.6/ptodata/1/aa/5B.COMB.pep.*
3: /cgn2.6/ptodata/1/aa/6A.COMB.pep.*
4: /cgn2.6/ptodata/1/aa/6B.COMB.pep.*
5: /cgn2.6/ptodata/1/aa/PCYTUS.COMB.pep.*
6: /cgn2.6/ptodata/1/aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	1.4	253	4 US-09-149-476-576	Sequence 576, App
2	7	1.4	345	4 US-09-194-905-12	Sequence 12, App1
3	7	1.4	589	2 US-08-937-540-6	Sequence 6, App1
4	7	1.4	590	4 US-09-398-395A-54	Sequence 54, App1
5	7	1.4	749	2 US-08-937-931-8	Sequence 8, App1
6	7	1.4	749	4 US-09-285-502-8	Sequence 8, App1
7	7	1.4	749	4 US-09-709-126-8	Sequence 8, App1
8	7	1.4	749	4 US-08-486-099-103	Sequence 103, App1
9	7	1.4	856	3 US-08-484-223B-103	Sequence 103, App
10	7	1.4	856	3 US-08-919-597-103	Sequence 103, App
11	7	1.4	856	3 US-08-475-668A-103	Sequence 103, App
12	7	1.4	856	3 US-08-485-551A-103	Sequence 103, App
13	7	1.4	856	3 US-08-471-913A-103	Sequence 103, App
14	7	1.4	856	4 US-08-485-264A-103	Sequence 103, App
15	7	1.4	856	4 US-08-474-349A-103	Sequence 103, App
16	7	1.4	856	1 US-08-220-151-10	Sequence 10, App1
17	7	1.4	857	1 US-08-413-118-10	Sequence 10, App1
18	7	1.4	857	1 US-08-804-439A-18	Sequence 18, App1
19	7	1.4	857	3 US-08-360-107A-113	Sequence 113, App
20	7	1.4	857	3 US-08-473-446-10	Sequence 10, App1
21	7	1.4	857	4 US-08-720-329-18	Sequence 18, App1
22	7	1.4	857	4 US-09-293-322C-11	Sequence 11, App1
23	6	1.2	9	4 US-09-602-999A-240	Sequence 240, App
24	6	1.2	16	4 US-09-500-124-240	Sequence 15, App1
25	6	1.2	27	3 US-08-749-816-15	Sequence 64, App1
26	6	1.2	37	3 US-08-651-136C-64	
27					

ALIGNMENTS

RESULT 1	US-09-149-476-576	US-09-229-911A-64
Sequence 576, Application	US-09-149-476	US-09-229-911A-64
Patent No. 6420526		
GENERAL INFORMATION:		
APPLICANT: Rosen et al.		
TITLE OF INVENTION: 186 Human Secreted proteins		
FILE REFERENCE: P2002P1		
CURRENT APPLICATION NUMBER: US-09-149-476		
CURRENT FILING DATE: 1998-09-08		
EARLIER APPLICATION NUMBER: PCT/US98/04493		
EARLIER FILING DATE: 1998-03-06		
EARLIER APPLICATION NUMBER: 60/040,162		
EARLIER FILING DATE: 1997-03-07		
EARLIER APPLICATION NUMBER: 60/040,333		
EARLIER FILING DATE: 1997-03-07		
EARLIER APPLICATION NUMBER: 60/038,621		
EARLIER FILING DATE: 1997-03-07		
EARLIER APPLICATION NUMBER: 60/040,626		
EARLIER FILING DATE: 1997-03-07		
EARLIER APPLICATION NUMBER: 60/040,334		
EARLIER FILING DATE: 1997-03-07		
EARLIER APPLICATION NUMBER: 60/040,336		
EARLIER FILING DATE: 1997-03-07		
EARLIER APPLICATION NUMBER: 60/040,163		
EARLIER FILING DATE: 1997-03-07		
EARLIER APPLICATION NUMBER: 60/047,600		
EARLIER FILING DATE: 1997-05-23		
EARLIER APPLICATION NUMBER: 60/047,615		
EARLIER FILING DATE: 1997-05-23		
EARLIER APPLICATION NUMBER: 60/047,597		
EARLIER FILING DATE: 1997-05-23		
EARLIER APPLICATION NUMBER: 60/047,502		
EARLIER FILING DATE: 1997-05-23		
EARLIER APPLICATION NUMBER: 60/047,633		
EARLIER FILING DATE: 1997-05-23		
EARLIER APPLICATION NUMBER: 60/047,583		
EARLIER FILING DATE: 1997-05-23		
EARLIER APPLICATION NUMBER: 60/047,617		
EARLIER FILING DATE: 1997-05-23		
EARLIER APPLICATION NUMBER: 60/047,618		
EARLIER FILING DATE: 1997-05-23		
EARLIER APPLICATION NUMBER: 60/047,503		
EARLIER FILING DATE: 1997-05-23		
EARLIER APPLICATION NUMBER: 60/047,592		
EARLIER FILING DATE: 1997-05-23		
EARLIER APPLICATION NUMBER: 60/047,581		
EARLIER FILING DATE: 1997-05-23		
EARLIER APPLICATION NUMBER: 60/047,584		
EARLIER FILING DATE: 1997-05-23		

28	6	1.2	37	4	US-09-229-911A-64	Sequence 64, App1
29	6	1.2	59	4	US-09-384-302A-16	Sequence 16, App1
30	6	1.2	62	4	US-09-134-001C-4865	Sequence 4865, App
31	6	1.2	63	4	US-08-469-260A-431	Sequence 431, App
32	6	1.2	72	4	US-08-858-207A-443	Sequence 443, App
33	6	1.2	85	2	US-08-480-229C-7	Sequence 7, App1
34	6	1.2	80	4	US-08-659-235C-17	Sequence 17, App1
35	6	1.2	95	2	US-09-384-302A-17	Sequence 17, App1
36	6	1.2	93	3	US-09-208-804-3	Sequence 3, App1
37	6	1.2	93	3	US-08-801-743-3	Sequence 3, App1
38	6	1.2	102	4	US-09-199-637A-91	Sequence 91, App1
39	6	1.2	104	1	US-08-111-939A-26	Sequence 26, App1
40	6	1.2	106	4	US-09-199-637A-59	Sequence 59, App1
41	6	1.2	108	2	US-08-162-402B-27	Sequence 27, App1
42	6	1.2	127	4	US-09-160-246-16	Sequence 16, App1
43	6	1.2	134	4	US-09-201-226-2	Sequence 2, App1
44	6	1.2	134	4	US-09-201-227A-16	Sequence 16, App1
45	6	1.2	135	4	US-09-134-001C-4905	Sequence 4905, App

Mon Dec 2 12:19:45 2002

us-09-868-760-7.oli.rai

Page 2

[illegible]

1	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/056,894
2	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/056,894
3	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/056,911
4	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/056,911
5	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/056,936
6	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/056,936
7	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/056,874
8	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/056,874
9	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/056,910
10	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/056,864
11	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/056,631
12	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/056,845
13	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/056,892
14	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/057,761
15	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/047,595
16	EARLIER	FILING DATE:	1997-05-23	EARLIER	APPLICATION NUMBER:	60/047,599
17	EARLIER	FILING DATE:	1997-05-23	EARLIER	APPLICATION NUMBER:	60/047,588
18	EARLIER	FILING DATE:	1997-05-23	EARLIER	APPLICATION NUMBER:	60/047,585
19	EARLIER	FILING DATE:	1997-05-23	EARLIER	APPLICATION NUMBER:	60/047,586
20	EARLIER	FILING DATE:	1997-05-23	EARLIER	APPLICATION NUMBER:	60/047,590
21	EARLIER	FILING DATE:	1997-05-23	EARLIER	APPLICATION NUMBER:	60/047,594
22	EARLIER	FILING DATE:	1997-05-23	EARLIER	APPLICATION NUMBER:	60/047,589
23	EARLIER	FILING DATE:	1997-05-23	EARLIER	APPLICATION NUMBER:	60/047,593
24	EARLIER	FILING DATE:	1997-05-22	EARLIER	APPLICATION NUMBER:	60/047,614
25	EARLIER	FILING DATE:	1997-05-23	EARLIER	APPLICATION NUMBER:	60/043,578
26	EARLIER	FILING DATE:	1997-04-11	EARLIER	APPLICATION NUMBER:	60/043,576
27	EARLIER	FILING DATE:	1997-04-11	EARLIER	APPLICATION NUMBER:	60/047,501
28	EARLIER	FILING DATE:	1997-05-23	EARLIER	APPLICATION NUMBER:	60/043,670
29	EARLIER	FILING DATE:	1997-04-11	EARLIER	APPLICATION NUMBER:	60/056,632
30	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/056,654
31	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/056,876
32	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/056,881
33	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/056,909
34	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/056,875
35	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/056,862
36	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/056,887
37	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/056,908
38	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/048,964
39	EARLIER	FILING DATE:	1997-05-06	EARLIER	APPLICATION NUMBER:	60/057,650
40	EARLIER	FILING DATE:	1997-09-05	EARLIER	APPLICATION NUMBER:	60/056,884
41	EARLIER	FILING DATE:	1997-08-22	EARLIER	APPLICATION NUMBER:	60/056,894

```

; EARLIER APPLICATION NUMBER: 60/057,669
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/049,610
; EARLIER FILING DATE: 1997-06-13
; EARLIER APPLICATION NUMBER: 60/061,060
; EARLIER FILING DATE: 1997-10-02

Query Match          1.4%; Score 7; DB 4; Length 253;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 301 HGYNES 307
DB 169 HGYNES 175

RESULT 2
US-09-194-905-12
; Sequence 12, Application US/09194905
; Patent No. 6306627
; GENERAL INFORMATION:
; APPLICANT: DECKER, Heinrich
; TITLE OF INVENTION: ISOLATION OF THE BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: PSEUDO-OLIGOSACCHARIDES FROM STREPTOMYCES GLAUCESCENS
; TITLE OF INVENTION: GLA.O AND THEIR USE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/194,905
; FILING DATE: 29-JUL-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/EP97/02826
; FILING DATE: 30-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 19622783.6
; FILING DATE: 07-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Granados, Patricia D.
; REGISTRATION NUMBER: 33,683
; REFERENCE/DOCKET NUMBER: 026083/0193
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 345 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-194-905-12

Query Match          1.4%; Score 7; DB 4; Length 345;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 383 QTPGST 389
DB 93 QTPGST 99

; EARLIER APPLICATION NUMBER: 60/057,669
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/049,610
; EARLIER FILING DATE: 1997-06-13
; EARLIER APPLICATION NUMBER: 60/061,060
; EARLIER FILING DATE: 1997-10-02

Query Match          1.4%; Score 7; DB 4; Length 253;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 301 HGYNES 307
DB 169 HGYNES 175

RESULT 3
US-08-937-540-6
; Sequence 6, Application US/08937540
; Patent No. 5891697
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Wise, Mitchell J
; APPLICANT: Savage, Thomas J
; APPLICANT: Katsuta, Eva J
; TITLE OF INVENTION: Monoterpene Synthases from Common Sage
; TITLE OF INVENTION: (Salvia officinalis)
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON & KINDNESS
; STREET: 1420 FIFTH AVENUE
; CITY: SEATTLE
; STATE: WASHINGTON
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,540
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelton, Dennis K
; REGISTRATION NUMBER: 26,997
; REFERENCE/DOCKET NUMBER: WSUR111254
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206 695 1718
; TELEFAX: 206 224 0779
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 589 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-937-540-6

Query Match          1.4%; Score 7; DB 2; Length 589;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 397 GPNVSD 403
DB 156 GPNVSD 162

RESULT 4
US-09-398-395A-54
; Sequence 54, Application US/09398395A
; Patent No. 6468772
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: NO. 6468772L, Joseph P.
; APPLICANT: Starks, Courtney M.
; APPLICANT: Manna, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/398,395A
; CURRENT FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: 60/100,993
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/130,628
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/150,262
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: fastseq for Windows Version 3.0
```

SEQ ID NO 54
LENGTH: 590
TYPE: prt
ORGANISM: Salvia officinalis
US-09-398-395A-54

Query Match 1.4%; Score 7; DB 4; Length 590;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 397 GENVED 403
Db 156 GENVED 162

RESULT 5
US-08-937-931-8
Sequence 8, Application US/08937931
Patent No. 5935792
GENERAL INFORMATION:
APPLICANT: Rubin, Gerald M.
APPLICANT: Pan, Duojia
APPLICANT: Rooke, Jenny
APPLICANT: Yavari, Reza
APPLICANT: Xu, Tian
TITLE OF INVENTION: KUZ: A No. 5935792el Family of Metalloproteases
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/937,931
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: B97-081
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 749 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-937-931-8
Query Match 1.4%; Score 7; DB 2; Length 749;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 AMPEKHA 117
Db 196 AMPEKHA 202

RESULT 6
US-09-285-502-8
Sequence 8, Application US/09285502
Patent No. 6190876
GENERAL INFORMATION:

APPLICANT: Rubin, Gerald M.
APPLICANT: Pan, Duojia
APPLICANT: Rooke, Jenny
APPLICANT: Yavari, Reza
APPLICANT: Xu, Tian
TITLE OF INVENTION: KUZ: A No. 6190876el Family of Metalloproteases
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/285,502
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/937,931
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: B97-081
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 749 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-285-502-8

Query Match 1.4%; Score 7; DB 4; Length 749;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 AMPEKHA 117
Db 196 AMPEKHA 202

RESULT 7
US-09-709-126-8
Sequence 8, Application US/09709126
Patent No. 6319704
GENERAL INFORMATION:
APPLICANT: Rubin, Gerald M.
APPLICANT: Pan, Duojia
APPLICANT: Rooke, Jenny
APPLICANT: Yavari, Reza
APPLICANT: Xu, Tian
TITLE OF INVENTION: KUZ: A No. 6319704el Family of Metalloproteases
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/709,126
APPLICATION NUMBER: US/09/709,126
FILING DATE: 08-NO. 6319704-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/285,502
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 749 amino acids
TYPE: amino acid
STRANDEDNESS: Single
MOLECULE TYPE: Peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-709-126-8

Query Match 1.4%; Score 7; DB 4; Length 749;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 AHPEKHA 117
Db 196 AHPEKHA 202

RESULT 8
US-09-871-385A-8
Sequence 8, Application US/09871385A
Patent No. 6399350
GENERAL INFORMATION:
APPLICANT: Pan, Gerald M.
Pan, Duoqia
Rooke, Jenny
Yavari, Reza
Xu, Tian
TITLE OF INVENTION: K12: A No. 6399350el Family of Metalloproteases
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/871,385A
FILING DATE: 31-May-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/709,126
FILING DATE: 08-NO. 6399350-2000
APPLICATION NUMBER: 09/285,502
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: B97-081
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 749 amino acids
TYPE: amino acid
STRANDEDNESS: Single
MOLECULE TYPE: Peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-871-385A-8

Query Match 1.4%; Score 7; DB 4; Length 749;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 AHPEKHA 117
Db 196 AHPEKHA 202

RESULT 9
US-08-486-099-103
Sequence 103 Application US/08486099
Patent No. 6013263
GENERAL INFORMATION:
APPLICANT: Bolognesi, Daniel P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Pelletway, Stephen R.
TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HEPATITIS
B VIRUS TRANSMISSION
TITLE OF INVENTION: B VIRUS TRANSMISSION
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESS: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,099
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-031
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 856 amino acids
TYPE: amino acid
STRANDEDNESS:
MOLECULE TYPE: Protein
US-08-486-099-103

Query Match 1.4%; Score 7; DB 3; Length 856;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 188 PETAAL 194
DB 842 PETAAL 848

RESULT 10

US-08-484-223B-103

Sequence 103, Application US/08484223B
Patent No. 6020459

GENERAL INFORMATION:

APPLICANT: Bolognesi, Dani P.
APPLICANT: Matthews, Thomas J.

APPLICANT: Wild, Carl T.

APPLICANT: Barney, Shawn O.

APPLICANT: Lambert, Dennis M.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

US-08-484-223B-103

Query Match 1.4%; Score 7; DB 3; Length 856;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 188 PETAAL 194
DB 842 PETAAL 848

RESULT 11

US-08-919-597-103

Sequence 103, Application US/08919597
Patent No. 6054265

GENERAL INFORMATION:

APPLICANT: Bolognesi, Dani P.

APPLICANT: Matthews, Thomas J.

APPLICANT: Wild, Carl T.

APPLICANT: Barney, Shawn O.

APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
APPLICANT: Langlois, Alphonse J.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
TRANSMISSION
TITLE OF INVENTION: TRANSMISSION
NUMBER OF SEQUENCES: 273
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/919,597
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/470,896
FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-020
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 856 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-919-597-103

Query Match 1.4%; Score 7; DB 3; Length 856;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 188 PETAAL 194
DB 842 PETAAL 848

RESULT 12

US-08-475-668A-103

Sequence 103, Application US/08475668A
Patent No. 6050065

GENERAL INFORMATION:

APPLICANT: Barney, Shawn O.

APPLICANT: Lambert, Dennis M.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

APPLICANT: Langlois, Alphonse J.

APPLICANT: Petteway, Stephen R.

US-08-475-668A-103

Sequence 103, Application US/08475668A
Patent No. 6050065

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 188 PETAAL 194
DB 842 PETAAL 848

RESULT 12

US-08-919-597-103

Sequence 103, Application US/08919597
Patent No. 6054265

GENERAL INFORMATION:

APPLICANT: Bolognesi, Dani P.

APPLICANT: Matthews, Thomas J.

APPLICANT: Wild, Carl T.

APPLICANT: Barney, Shawn O.

APPLICANT: Lambert, Dennis M.

US-08-475-668A-103

Sequence 103, Application US/08475668A
Patent No. 6050065

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 188 PETAAL 194
DB 842 PETAAL 848

RESULT 12

US-08-919-597-103

Sequence 103, Application US/08919597
Patent No. 6054265

GENERAL INFORMATION:

APPLICANT: Bolognesi, Dani P.

APPLICANT: Matthews, Thomas J.

APPLICANT: Wild, Carl T.

APPLICANT: Barney, Shawn O.

APPLICANT: Lambert, Dennis M.

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,668A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-026
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 856 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-475-668A-103

Query Match 1.4%; Score 7; DB 3; Length 856;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 188 PETAAL 194
Db 842 PETAAL 848

RESULT 13
US-08-485-551A-103
Sequence 103, Application US/08485551A
Patent No. 6068973
GENERAL INFORMATION:
APPLICANT: Bolognesi, Dani P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
APPLICANT: Langlois, Alphonse J.
TITLE OF INVENTION: METHODS FOR INHIBITION OF MEMBRANE
TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
TITLE OF INVENTION: TRANSMISSION
NUMBER OF SEQUENCES: 211
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,551A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-023
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 856 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-485-551A-103

Query Match 1.4%; Score 7; DB 3; Length 856;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 188 PETAAL 194
Db 842 PETAAL 848

RESULT 14
US-08-471-913A-103
Sequence 103, Application US/08471913A
Patent No. 6093794
GENERAL INFORMATION:
APPLICANT: Bolognesi, Dani P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
APPLICANT: Langlois, Alphonse J.
TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING EPSTEIN-BARR VIRUS
TITLE OF INVENTION: TRANSMISSION
NUMBER OF SEQUENCES: 214
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,913A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-030
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 856 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-471-913A-103

Query Match 1.4%; Score 7; DB 3; Length 856;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 188 PETAAL 194
Db 842 PETAAL 848

DB 842 PETAAL 848

```

RESULT 15
US-08-485-264A-103
; Sequence 103, Application US/08485264A
; Patent No. 6228983
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Peteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING
; TITLE OF INVENTION: RESPIRATORY SYNCYTIAL VIRUS TRANSMISSION
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,264A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 433
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 969-9090
; TELEFAX: (212) 969-9741/8864
; TELEX: 66141 DENNIE
; INFORMATION FOR SEQ ID NO: 103:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 856 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-485-264A-103

Query Match 1.4%: Score 7; DB 4: Length 856;
Best Local Similarity 100.0%: Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 188 PETAAL 194
DB 842 PETAAL 848

```

Search completed: November 29, 2002, 12:47:33
 Job time : 25 secs

PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 40510
LENGTH: 18
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC011159.3
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
US-09-864-761-40510
Query Match 1.6%; Score 8; DB 10; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CY 23 HDDHDDDD 30
DB 9 HDDHDDDD 16
RESULT 2
US-08-910-386A-3
Sequence 3, Application US/08910386A
Patent No. US20020092041A1
GENERAL INFORMATION:
APPLICANT: Ronald, Pamela C.
APPLICANT: Wang, Guo-Liang
APPLICANT: Song, Wen-Yuang
APPLICANT: Hulbert, Scott
APPLICANT: Richter, Todd
TITLE OF INVENTION: Procedures and Materials for Confering
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,386A
FILING DATE: 13-AUG-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 023070-058950US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1445 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-910-386A-3
Query Match 1.6%; Score 8; DB 8; Length 1445;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CY 273 CDNLGARY 280
DB 1372 CDNLGARY 1379
RESULT 3
US-09-764-868-863
Sequence 863, Application US/09764868
Patent No. US20020168711A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT232
CURRENT APPLICATION NUMBER: US/09/764,868
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 1510
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 863
LENGTH: 192
TYPE: PRT
ORGANISM: Homo sapiens
US-09-764-868-863
Query Match 1.4%; Score 7; DB 9; Length 192;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CY 470 LMARSLV 476
DB 174 LMARSLV 180
RESULT 4
US-09-764-868-1175
Sequence 1175, Application US/09764868
Patent No. US20020168711A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT232
CURRENT APPLICATION NUMBER: US/09/764,868
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 1510
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1175
LENGTH: 222
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (190)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (194)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (195)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (205)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-1175

```
Query Match 1.4%: Score 7; DB 9; Length 222;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 LNARSLV 476
DB 174 LNARSLV 180

RESULT 5
US-09-815-242-13488
; Sequence 13488, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes In
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13488
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13488

Query Match 1.4%: Score 7; DB 10; Length 328;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 LNARSLV 476
DB 50 LNARSLV 56

RESULT 6
US-09-815-242-13556
; Sequence 13556, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes In

; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13556
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13556

Query Match 1.4%: Score 7; DB 10; Length 328;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 LNARSLV 476
DB 50 LNARSLV 56

RESULT 7
US-09-802-472B-2
; Sequence 2, Application US/09802472B
; Patent No. US2002010353A1
; GENERAL INFORMATION:
; APPLICANT: EINAT, Paz
; APPLICANT: SKALITER, Ramt
; APPLICANT: FEINSTEIN, Elena
; TITLE OF INVENTION: SEQUENCES CHARACTERISTIC OF HYPOXIA-REGULATED GENE TRANSCRIPTS
; FILE REFERENCE: EINAT-4.1C
; CURRENT APPLICATION NUMBER: US/09/802,472B
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: US 09/383,096
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: US 09/138,109
; PRIOR FILING DATE: 1998-08-21
; PRIOR APPLICATION NUMBER: US 60/098,158
; PRIOR FILING DATE: 1998-08-27
; PRIOR APPLICATION NUMBER: US 60/132,684
; PRIOR FILING DATE: 1999-05-05
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-802-472B-2

Query Match 1.4%: Score 7; DB 10; Length 437;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 301 HGVNES 307
DB 354 HGVNES 360

RESULT 8
```

US-09-815-242-5114
; Sequence 5114, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: EPIRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5114
; LENGTH: 531
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5114
Query Match 1.4%; Score 7; DB 10; Length 531;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 119 PGDLGDL 125
Db 278 PGDLGDL 284

US-09-887-586A-54
Query Match 1.4%; Score 7; DB 10; Length 590;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 397 GFNVSED 403
Db 156 GFNVSED 162

US-09-903-012-54
; Sequence 54, Application US/09903012
; Patent No. US20020094557A1
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: Starks, Courtney M.
; APPLICANT: Manna, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/903,012
; PRIOR FILING DATE: 2001-07-11
; PRIOR FILING DATE: 1999-09-17
; PRIOR FILING DATE: 1998-09-18
; PRIOR FILING DATE: 1998-09-18
; PRIOR FILING DATE: 1999-04-22
; PRIOR FILING DATE: 1999-04-22
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 54
; LENGTH: 590
; TYPE: PRT
; ORGANISM: Salvia officinalis
US-09-903-012-54
Query Match 1.4%; Score 7; DB 10; Length 590;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 397 GFNVSED 403
Db 156 GFNVSED 162

US-10-052-586-560
; Sequence 560, Application US/10052586
; Patent No. US20020127584A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Wetlands, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C1
; CURRENT APPLICATION NUMBER: US/10/052,586
; PRIOR FILING DATE: 2002-01-15
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266

us-09-868-760-7.oli.rabb

1	PRIOR FILING DATE:	1997-09-18
2	PRIOR APPLICATION NUMBER:	60/0662250
3	PRIOR FILING DATE:	1997-10-17
4	PRIOR APPLICATION NUMBER:	60/063120
5	PRIOR FILING DATE:	1997-10-24
6	PRIOR APPLICATION NUMBER:	60/063121
7	PRIOR FILING DATE:	1997-10-24
8	PRIOR APPLICATION NUMBER:	60/063486
9	PRIOR FILING DATE:	1997-10-21
10	PRIOR APPLICATION NUMBER:	60/063540
11	PRIOR FILING DATE:	1997-10-28
12	PRIOR APPLICATION NUMBER:	60/063541
13	PRIOR FILING DATE:	1997-10-28
14	PRIOR APPLICATION NUMBER:	60/063544
15	PRIOR FILING DATE:	1997-10-28
16	PRIOR APPLICATION NUMBER:	60/063564
17	PRIOR FILING DATE:	1997-10-28
18	PRIOR APPLICATION NUMBER:	60/063734
19	PRIOR FILING DATE:	1997-10-29
20	PRIOR APPLICATION NUMBER:	60/063870
21	PRIOR FILING DATE:	1997-10-31
22	PRIOR APPLICATION NUMBER:	60/064103
23	PRIOR FILING DATE:	1997-10-31
24	PRIOR APPLICATION NUMBER:	60/065311
25	PRIOR FILING DATE:	1997-11-07
26	PRIOR APPLICATION NUMBER:	60/0656120
27	PRIOR FILING DATE:	1997-11-24
28	PRIOR APPLICATION NUMBER:	60/066466
29	PRIOR FILING DATE:	1997-11-24
30	PRIOR APPLICATION NUMBER:	60/066772
31	PRIOR FILING DATE:	1997-12-11
32	PRIOR APPLICATION NUMBER:	60/069335
33	PRIOR FILING DATE:	1997-12-11
34	PRIOR APPLICATION NUMBER:	60/069425
35	PRIOR FILING DATE:	1997-12-12
36	PRIOR APPLICATION NUMBER:	60/069670
37	PRIOR FILING DATE:	1997-12-17
38	PRIOR APPLICATION NUMBER:	60/068017
39	PRIOR FILING DATE:	1997-12-18
40	PRIOR APPLICATION NUMBER:	60/077450
41	PRIOR FILING DATE:	1998-03-10
42	PRIOR APPLICATION NUMBER:	60/077632
43	PRIOR FILING DATE:	1998-03-11
44	PRIOR APPLICATION NUMBER:	60/077649
45	PRIOR FILING DATE:	1998-03-11
46	PRIOR APPLICATION NUMBER:	60/078886
47	PRIOR FILING DATE:	1998-03-20
48	PRIOR APPLICATION NUMBER:	60/078939
49	PRIOR FILING DATE:	1998-03-20
50	PRIOR APPLICATION NUMBER:	60/079664
51	PRIOR FILING DATE:	1998-03-27
52	PRIOR APPLICATION NUMBER:	60/079786
53	PRIOR FILING DATE:	1998-03-27
54	PRIOR APPLICATION NUMBER:	60/080107
55	PRIOR FILING DATE:	1998-03-31
56	PRIOR APPLICATION NUMBER:	60/080194
57	PRIOR FILING DATE:	1998-03-31
58	PRIOR APPLICATION NUMBER:	60/080327
59	PRIOR FILING DATE:	1998-04-01
60	PRIOR APPLICATION NUMBER:	60/080333
61	PRIOR FILING DATE:	1998-04-01
62	PRIOR APPLICATION NUMBER:	60/081049
63	PRIOR FILING DATE:	1998-04-08
64	PRIOR APPLICATION NUMBER:	60/081070
65	PRIOR FILING DATE:	1998-04-08
66	PRIOR APPLICATION NUMBER:	60/081195
67	PRIOR FILING DATE:	1998-04-09
68	PRIOR APPLICATION NUMBER:	60/081838
69	PRIOR FILING DATE:	1998-04-15
70	PRIOR APPLICATION NUMBER:	60/082568
71	PRIOR FILING DATE:	1998-04-21
72	PRIOR APPLICATION NUMBER:	60/082569
73	PRIOR FILING DATE:	1998-04-21
74	PRIOR APPLICATION NUMBER:	60/082569

```

1  PRIOR APPLICATION NUMBER: 60/082704
2  PRIOR FILING DATE: 1998-04-22
3  PRIOR APPLICATION NUMBER: 60/082757
4  PRIOR FILING DATE: 1998-04-22
5  PRIOR APPLICATION NUMBER: 60/083352
6  PRIOR FILING DATE: 1998-04-28
7  PRIOR APPLICATION NUMBER: 60/083495
8  PRIOR FILING DATE: 1998-04-29
9  PRIOR APPLICATION NUMBER: 60/083496
10 PRIOR FILING DATE: 1998-04-29
11 PRIOR APPLICATION NUMBER: 60/083499
12 PRIOR FILING DATE: 1998-04-29
13 PRIOR APPLICATION NUMBER: 60/083559
14 PRIOR FILING DATE: 1998-04-29
15 PRIOR APPLICATION NUMBER: 60/084366
16 PRIOR FILING DATE: 1998-05-05
17 PRIOR APPLICATION NUMBER: 60/084414
18 PRIOR FILING DATE: 1998-05-06
19 PRIOR APPLICATION NUMBER: 60/084639
20 PRIOR FILING DATE: 1998-05-07
21 PRIOR APPLICATION NUMBER: 60/084640
22 PRIOR FILING DATE: 1998-05-07
23 PRIOR APPLICATION NUMBER: 60/084643
24 PRIOR FILING DATE: 1998-05-07
25 PRIOR APPLICATION NUMBER: 60/085573
26 PRIOR FILING DATE: 1998-05-15
27 PRIOR APPLICATION NUMBER: 60/085579
28 PRIOR FILING DATE: 1998-05-15
29 PRIOR APPLICATION NUMBER: 60/085580
30 PRIOR FILING DATE: 1998-05-15
31 PRIOR APPLICATION NUMBER: 60/085582
32 PRIOR FILING DATE: 1998-05-15
33 PRIOR APPLICATION NUMBER: 60/085700
34 PRIOR FILING DATE: 1998-05-15
35 PRIOR APPLICATION NUMBER: 60/086023
36 PRIOR FILING DATE: 1998-05-18
37 PRIOR APPLICATION NUMBER: 60/086392
38 PRIOR FILING DATE: 1998-05-22
39 PRIOR APPLICATION NUMBER: 60/086486
40 PRIOR FILING DATE: 1998-05-22
41 PRIOR APPLICATION NUMBER: 60/087098
42 PRIOR FILING DATE: 1998-05-28
43 PRIOR APPLICATION NUMBER: 60/087208
44 PRIOR FILING DATE: 1998-05-28
45 PRIOR APPLICATION NUMBER: 60/087609
46 PRIOR FILING DATE: 1998-06-02
47 PRIOR APPLICATION NUMBER: 60/087759
48 PRIOR FILING DATE: 1998-06-02
49 PRIOR APPLICATION NUMBER: 60/087827
50 PRIOR FILING DATE: 1998-06-03
51 PRIOR APPLICATION NUMBER: 60/088025
52 PRIOR FILING DATE: 1998-06-04
53 PRIOR APPLICATION NUMBER: 60/088028
54 PRIOR FILING DATE: 1998-06-04
55 PRIOR APPLICATION NUMBER: 60/088039
56 PRIOR FILING DATE: 1998-06-04
57 PRIOR APPLICATION NUMBER: 60/088033
58 PRIOR FILING DATE: 1998-06-04
59 PRIOR APPLICATION NUMBER: 60/088167
60 PRIOR FILING DATE: 1998-06-05
61 PRIOR APPLICATION NUMBER: 60/088202
62 PRIOR FILING DATE: 1998-06-05
63 PRIOR APPLICATION NUMBER: 60/088212
64 PRIOR FILING DATE: 1998-06-05
65 PRIOR APPLICATION NUMBER: 60/088327
66 PRIOR FILING DATE: 1998-06-05
67 PRIOR APPLICATION NUMBER: 60/088326
68 PRIOR FILING DATE: 1998-06-05
69 PRIOR APPLICATION NUMBER: 60/088655
70 PRIOR FILING DATE: 1998-06-09
71 PRIOR APPLICATION NUMBER: 60/088722
72 PRIOR FILING DATE: 1998-06-10
73 PRIOR APPLICATION NUMBER: 60/088773

```

PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088740
 PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088811
 PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088824
 PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088825
 PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088826
 PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088861
 PRIOR FILING DATE: 1998-06-11
 PRIOR APPLICATION NUMBER: 60/088863
 PRIOR FILING DATE: 1998-06-11
 PRIOR APPLICATION NUMBER: 60/088876
 PRIOR FILING DATE: 1998-06-11
 PRIOR APPLICATION NUMBER: 60/089090
 PRIOR FILING DATE: 1998-06-12
 PRIOR APPLICATION NUMBER: 60/089105
 PRIOR FILING DATE: 1998-06-12
 PRIOR APPLICATION NUMBER: 60/089512
 PRIOR FILING DATE: 1998-06-16
 PRIOR APPLICATION NUMBER: 60/089514
 PRIOR FILING DATE: 1998-06-16
 PRIOR APPLICATION NUMBER: 60/089538
 PRIOR FILING DATE: 1998-06-17
 PRIOR APPLICATION NUMBER: 60/089598
 PRIOR FILING DATE: 1998-06-17
 PRIOR APPLICATION NUMBER: 60/089653
 PRIOR FILING DATE: 1998-06-17
 PRIOR APPLICATION NUMBER: 60/089908

Query Match 1.4% Score 7: DB 12: Length 692;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 7: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 189 ETAAALH 195
 DB 123 ETAAALH 129

RESULT 12
 US-09-871-388-8
 Sequence 8, Application US/09871388
 Patent No. US20020127621A1
 GENERAL INFORMATION:
 APPLICANT: Rubin, Gerald M.
 Pan, Duojia
 Rooke, Jenny
 Yavari, Reza
 Xu, Tian
 TITLE OF INVENTION: KUZ: A NO. US20020127621A1el Family of Metalloproteases
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: SCIENCE & TECHNOLOGY LAM GROUP
 STREET: 268 BUSH STREET, SUITE 3200
 CITY: SAN FRANCISCO
 STATE: CALIFORNIA
 COUNTRY: USA
 ZIP: 94104
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/871,388
 FILING DATE: 31-May-2001
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/937,931
 FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
 NAME: OSMAN, RICHARD A
 REGISTRATION NUMBER: 36,627
 REFERENCE/DOCKET NUMBER: B97-081
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 343-4341
 TELEFAX: (415) 343-4342
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 749 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 SEQUENCE DESCRIPTION: SEQ ID NO: 8:
 US-09-871-388-8

Query Match 1.4% Score 7: DB 10: Length 749;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 7: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 APERKHA 117
 DB 196 APERKHA 202

RESULT 13
 US-09-815-242-13636
 Sequence 13636, Application US/09815242
 Patent No. US2002061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 Olsen, Karl L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel D.
 APPLICANT: Traxler, John J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 PRIOR FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 13636
 LENGTH: 810
 TYPE: PRT
 ORGANISM: Streptococcus pneumoniae
 US-09-815-242-13636

Query Match 1.4% Score 7: DB 10: Length 810;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 7: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 158 SMTILOG 164
 DB 374 SMTILOG 380

```
RESULT 14
US-09-881-752A-308
; Sequence 308, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleenhou, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Clemen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides in the
; FILE REFERENCE: 06132/041002
; CURRENT APPLICATION NUMBER: US/09/881,752A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 308
; LENGTH: 1797
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 85
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-881-752A-308

Query Match
Best Local Similarity 1.4%; Score 7; DB 10; Length 1797;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 313 INIFGDD 319
|||||
Db 276 INIFGDD 282

RESULT 15
US-09-815-242-11410
; Sequence 11410, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
```

```
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11410
; LENGTH: 2890
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11410

Query Match
Best Local Similarity 1.4%; Score 7; DB 10; Length 2890;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 313 INIFGDD 319
|||||
Db 1369 INIFGDD 1375

Search completed: November 29, 2002, 12:48:06
Job time : 19 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2002, 12:42:30 ; Search time 20 Seconds

(without alignments)
2388.940 Million cell updates/sec

Title: US-09-868-760-7

Perfect score: 497

Sequence: 1 DGEQCDNGONKDDHDDHDD.....MOGHEVESERVACVIGRA 497

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : PIR-73.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	1.8	79	2 C84308	hypothetical prote
2	8	1.6	131	2 T25070	hypothetical prote
3	8	1.6	149	2 A54530	eggshell protein -
4	8	1.6	362	2 T08816	choline kinase (EC
5	8	1.6	378	2 AC0722	probable cytochrom
6	8	1.6	421	1 RK0ALG	ribulose-bisphosph
7	8	1.6	421	2 S21984	ribulose-bisphosph
8	8	1.6	426	1 RKBHLC	ribulose-bisphosph
9	8	1.6	441	1 RKBHLC	ribulose-bisphosph
10	8	1.6	452	2 S47228	ribulose-bisphosph
11	8	1.6	452	2 S47229	ribulose-bisphosph
12	8	1.6	452	2 S47230	ribulose-bisphosph
13	8	1.6	452	2 S47231	ribulose-bisphosph
14	8	1.6	452	2 S47232	ribulose-bisphosph
15	8	1.6	452	2 S47233	ribulose-bisphosph
16	8	1.6	452	2 S47234	ribulose-bisphosph
17	8	1.6	452	2 S47235	ribulose-bisphosph
18	8	1.6	452	2 S47236	ribulose-bisphosph
19	8	1.6	452	2 S47237	ribulose-bisphosph
20	8	1.6	452	2 S47238	ribulose-bisphosph
21	8	1.6	452	2 S47239	ribulose-bisphosph
22	8	1.6	452	2 S47240	ribulose-bisphosph
23	8	1.6	452	2 S47241	ribulose-bisphosph
24	8	1.6	452	2 S47242	ribulose-bisphosph
25	8	1.6	452	2 S47243	ribulose-bisphosph
26	8	1.6	452	2 S47244	ribulose-bisphosph
27	8	1.6	452	2 S47245	ribulose-bisphosph
28	8	1.6	452	2 S47246	ribulose-bisphosph
29	8	1.6	452	2 S47247	ribulose-bisphosph

30	8	1.6	469	1 RKBUL	ribulose-bisphosph
31	8	1.6	469	2 T01638	ribulose-bisphosph
32	8	1.6	469	2 T01640	ribulose-bisphosph
33	8	1.6	469	2 T01641	ribulose-bisphosph
34	8	1.6	469	2 T01644	ribulose-bisphosph
35	8	1.6	472	1 RKRRL2	ribulose-bisphosph
36	8	1.6	472	1 RKRRL	ribulose-bisphosph
37	8	1.6	472	1 RKRRL1	ribulose-bisphosph
38	8	1.6	472	2 T01642	ribulose-bisphosph
39	8	1.6	473	2 S34316	ribulose-bisphosph
40	8	1.6	473	2 S18315	ribulose-bisphosph
41	8	1.6	474	1 RKAALC	ribulose-bisphosph
42	8	1.6	475	1 RKRHLG	ribulose-bisphosph
43	8	1.6	475	1 RKRHLG	ribulose-bisphosph
44	8	1.6	475	1 RKRHLR	ribulose-bisphosph
45	8	1.6	475	1 RKSZLN	ribulose-bisphosph

ALIGNMENTS

RESULT 1
C84308
hypothetical protein Vng1546h [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence-revision 02-Feb-2001 #text-change 02-Feb-2001
C/Accession: C84308
R/ng, M.V.; Kennedy, S.P.; Mahabir, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky
; Leithausen, B.; Keller, K.; Gruen, R.; Dawson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
dun, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;
A>Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: C84308
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-79 <STO>
A:Cross-references: GB:AE004437; MID:910581035; PIDN:AG19831.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG1546H

Query Match
Best Local Similarity 1.8%; Score 9; DB 2; Length 79;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 125 LVDDDRGVY 133
DB 55 LVDDDRGVY 63

RESULT 2

T25070
hypothetical protein T21C9.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence-revision 15-Oct-1999 #text-change 15-Oct-1999
C/Accession: T25070
R/McMurray, A.
submitted to the EMBL Data Library, May 1996
A:Reference number: T25070
A:Accession: T25070
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-131 <WII>
A:Cross-references: EMBL:Z73098; PIDN:CAA97333.1; GSPDB:GN00023; CESP:T21C9.5
C:Genetics:
A:Experimental source: clone T21C9
A:Gene: CESP:T21C9.5
A:Map position: 5
A:Introns: 24/2; 68/3

Query Match
Best Local Similarity 1.6%; Score 8; DB 2; Length 131;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

	Matches	8 ;	Conservative	0 ;	Mismatches	0 ;	Indels	0 ;	Gaps	0 ;
QY	431	HGHDDAHD	438							
Db	124	HGHDDAHD	131							

RESULT 3
A54530
eggshell protein - fluke (Schistosoma mansoni) (fragment)
C:Species: Schistosoma mansoni
C:Date: 15-Oct-1994 #sequence_revision 15-Oct-1994 #text_change 15-Oct-1999
C:Accession: A54530
R:Johnson, K.S.; Taylor, D.W.; Cordingley, J.S.
Mol. Biochem. Parasitol. 22, 89-100, 1987
A:Title: Possible eggshell protein gene from Schistosoma mansoni.
A:Reference number: A54530; MID:87115624; PMID:3100949
A:Accession: A54530
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-149 <100>
A:Cross-references: GB:M15371; NID:J160976; PID:g160977
C:Keywords: egg shell; tandem repeat

Query Match	1.6%	Score 8	DB 2	Length 149
Best Local Similarity	100.0%	Pred. No. 2.7		
Matches	8	Conservative	0	Mismatches 0; Indels 0; Gaps 0
QY	13	DHDDHND	20	
Db	118	DHDDHND	125	

RESULT 4
T08816
chooline kinase (EC 2.7.1.32) CK2 - soybean
C:Species: Glycine max (soybean)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T08816
R:Dewey, R.E.; Monke, D.E.; Goode, J.H.
Plant Physiol. 110, 1197-1205, 1996
A:Title: Characterization of soybean choline kinase cDNAs and their expression in yeast
A:Reference number: 216478; MUID:97088716; PMID:8934624
A:Accession: T08816
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-362 <Dew>
A:Cross-references: EMBL:U43839; NID:g1438880; PIDD:AMC49375.1; PID:g1438881
A:Experimental source: strain Dare
C:Genetics;
A:Gene: CK2
C:Function:
A:Description: catalyzes phosphorylation of choline to O-phosphocholine in the presence
of keywords: phosphotransferase

```

Query Match      1.6%; Score 8; DB 2; Length 362;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      444 DLDGVIDD 451
          |||
Db       28 DLDGVIDD 35

RESULT 5
AC0722
probable cytochrome oxidase chain II [imported] - Salmonella enterica subsp. enterica serovar typhimurium
C:Species: Salmonella enterica subsp. enterica serovar typhi
A:Note: This species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AC07222
R:Perkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Parkhill, J.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

```

S. J. Moule, S. J. O'Garra, P. Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, A.; Little: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Paratyphi A
A:Reference number: AB0502; PMID:11677608
A:Accession: AC0722
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-378 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD05477.1; PID:g16502582; GSPDB:GN00176
C:Gene: STY1921
C:Superfamily: cytochrome d ubiquinol oxidase

Query Match 1.6%; Score 8; DB 2;
Best Local Similarity 100.0%; length 378;
Pred. No. 6, 5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	474	SLVIMQGG	481
Db	178	SLVIMQGG	185

RESULT 6
RKOALG
R:ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain - Aegilops squarrosa (fra
C:Species: Aegilops squarrosa
C:Date: 30-Sep-1992 #sequence_rev:130 30-Sep-1992 #text_change 23-Mar-2001
C:Accession: S17319
R:Ogihara, Y.; Terachi, T.; Sasahima, T.
Submitted to the EMBL Data Library, August 1991
A:Description: Molecular analysis of the hot spot region related to length mutations
region
A:Reference number: S17319
A:Accession: S17319
A:Molecule type: DNA
A:Residues: 1421<CGT>
A:Cross-references: EMBL:X62119, NID:g11324; PIDD:CAA4038.1; PID:g11325
C:Comment: In addition to Lys-146, another lysine, it is not certain which, may be th
C:Function:
A:Description: catalyzes the reaction of carbon dioxide with D-ribulose 1,5-bisphosph
C:Superfamily: ribulose-bisphosphate carboxylase large chain
C:Keywords: CALVIN cycle; carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase
E:146/279/Active site: Lys (ribulose-bisphosphate-binding) #status Predicted
E:146/Binding site: carbon dioxide (Lys) (covalent) (by Rubisco activase) #status pre
E:148/Binding site: magnesium (Asp) #status predicted

Query Match	1.6%;	Score 8;	DB 1;	Length 421;
Best Local Similarity	100.0%;	Pred. No. 7.2;		
Matches	8;	Conservative	0;	Mismatches
			0;	Indels
			0;	Gaps
			0;	
QY	315	IFGDDSVL	322	
Db	338	IFGDDSVL	345	

```

RESULT 7
S21984
Ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain - Aegilops crassa chlorop
C:Species: chloroplast Aegilops crassa
C:Date: 22-Nov-1993 #sequence_revision 26-Jul-1996 #text_change 15-Jun-2001
C:Accession: S21984
R:Ogihara, Y.; Terachi, T.; Sasahuma, T.
Submitted to the EMBL Data Library, August 1991
A:Description: Molecular analysis of the hot spot region related to length mutations
Region
A:Reference number: S17319
A:Accession: S21984
A:Molecule type: DNA
A:Residues: 1421 <CGT>
A:Cross-references: EMBL:X62118; NID:g11308; PIDD:CAA44032.1; PID:g11309
C:Genetics:
C:Gene: rbcL

```

```

A:Genome: chloroplast
C:Function:
A:Description: catalyzes the reaction of carbon dioxide with D-ribulose 1,5-bisphosphate
C:Superfamily: ribulose-bisphosphate carboxylase large chain
C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase; c
F:120/279/Active site: Lys (ribulose-bisphosphate-binding) #status predicted
F:148/Binding site: carbon dioxide (Lys) (covalent) (by Rubisco activase) #status predi
F:148/Binding site: magnesium (Asp) #status predicted
Query Match
Best Local Similarity 1.6%; Score 8; DB 2; Length 421;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 315 IFGDDSVL 322
Db 338 IFGDDSVL 345
|||||
RESULT 8
RHBLC
ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain precursor - barley chloropla
C:Species: chloroplast Hordeum vulgare (barley)
C>Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999
C:Accession: S08610
R:Zurawski, G.; Ciescy, M.T.; Brown, A.H.D.
Genetics 106:735-749, 1984
A:Title: The nature of nucleotide sequence divergence between barley and maize chloropla
A:Reference number: S07243
A:Accession: S08610
A:Molecule type: DNA
A:Residues: 1-426 <VAL>
A:Cross-references: EMBL:X00630; NID:911585; PIDN:CA5265.1; PID:911587
C:Comment: In addition to Lys-201, another lysine, it is not certain which, may be the s
C:Genetics:
A:Genome: chloroplast
C:Function:
A:Description: catalyzes the reaction of carbon dioxide with D-ribulose 1,5-bisphosphate
C:Superfamily: ribulose-bisphosphate carboxylase large chain
C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase; c
F:3-426/Product: ribulose-bisphosphate carboxylase large chain (fragment) #status predi
F:115/334/Active site: Lys (ribulose-bisphosphate-binding) #status predicted
F:201/Binding site: carbon dioxide (Lys) (covalent) (by Rubisco activase) #status predi
F:203/Binding site: magnesium (Asp) #status predicted
Query Match
Best Local Similarity 1.6%; Score 8; DB 1; Length 426;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 315 IFGDDSVL 322
Db 393 IFGDDSVL 400
|||||
RESULT 9
RHWLC
ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain - wheat chloroplast (fragme
C:Species: chloroplast Triticum aestivum (common wheat)
C>Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 23-Mar-2001
C:Accession: B33395; S17323
R:Hout, R.L.; Stults, J.T.; Mulligan, R.M.; Tolbert, N.E.
Proc. Natl. Acad. Sci. U.S.A. 86, 1855-1859, 1989
A:Title: Post-translational modifications in the large subunit of ribulose bisphosphate
A:Reference number: A33395; MUID:89184526; PMID:2928307
A:Accession: B33395
A:Molecule type: protein
A:Residues: 1-19 <HOU>
R:Ogihara, Y.; Terachi, T.; Sasahuma, T.
submitted to the EMBL Data Library, August 1991
A:Description: Molecular analysis of the hot spot region related to length mutations in
region.
A:Reference number: S17319
A:Accession: S17323
A:Molecule type: DNA

```

```

A:Residues: 20-441 <OCI>
A:Cross-references: EMBL:X62117; NID:912343; PIDN:CA44027.1; PID:912344
C:Comment: In addition to Lys-165, another lysine, it is not certain which, may be th
C:Genetics:
A:Gene: rbcL
A:Genome: chloroplast
C:Function:
A:Description: catalyzes the reaction of carbon dioxide with D-ribulose 1,5-bisphosph
C:Superfamily: ribulose-bisphosphate carboxylase large chain
C:Keywords: acetylated amino end; Calvin cycle; carbon dioxide fixation; carbon-carbo
F:1/Modified site: acetylated amino end (Pro) #status experimental
F:12/Modified site: N6,N6,N6-trimethyllysine (Lys) #status absent
F:139/298/Active site: Lys (ribulose-bisphosphate-binding) #status predicted
F:165/Binding site: carbon dioxide (Lys) (covalent) (by Rubisco activase) #status pre
F:167/Binding site: magnesium (Asp) #status predicted
Query Match
Best Local Similarity 1.6%; Score 8; DB 1; Length 441;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 315 IFGDDSVL 322
Db 357 IFGDDSVL 364
|||||
RESULT 10
S47228
ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain - Gallium corsicum chloro
C:Species: chloroplast Gallium corsicum
C>Date: 06-Jan-1995 #sequence_revision 12-May-1995 #text_change 22-Jun-1999
C:Accession: S47228; S39278
R:Manen, J.F.; Natali, A.
submitted to the EMBL Data Library, July 1994
A:Description: Comparison of the evolution of the rbcL sequence and of the atpB-rbcL
A:Accession number: S47221
A:Accession: S47228
A:Molecule type: DNA
A:Residues: 1-452 <MAN1>
A:Cross-references: EMBL:X81096
R:Manen, J.F.; Natali, A.; Ehrendorfer, F.
submitted to the EMBL Data Library, December 1993
A:Description: Phylogeny of Rubiaceae-Rubia inferred from the sequence of a cpDNA in
A:Reference number: S39270
A:Accession: S39278
A:Molecule type: DNA
A:Residues: 1-52, 'X', 54-57 <MAN2>
A:Cross-references: EMBL:X76463; NID:9434867; PIDN:CA54001.1; PID:9434868
C:Genetics:
A:Gene: rbcL
A:Genome: chloroplast
C:Function:
A:Description: catalyzes the reaction of carbon dioxide with D-ribulose 1,5-bisphosph
C:Superfamily: ribulose-bisphosphate carboxylase large chain
C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase
F:175/334/Active site: Lys (ribulose-bisphosphate-binding) #status predicted
F:201/Binding site: carbon dioxide (Lys) (covalent) (by Rubisco activase) #status pre
F:203/Binding site: magnesium (Asp) #status predicted
Query Match
Best Local Similarity 1.6%; Score 8; DB 2; Length 452;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 315 IFGDDSVL 322
Db 393 IFGDDSVL 400
|||||
RESULT 11
S47229
ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain - Gallium elongatum chloro
C:Species: chloroplast Gallium elongatum
C>Date: 06-Jan-1995 #sequence_revision 12-May-1995 #text_change 22-Jun-1999
C:Accession: S47229; S39279

```

R:Manen, J.F.; Natall, A.
 Submitted to the EMBL Data Library, July 1994
 C:Species: Chloroplast Gallium lucidum
 C:Date: 06-Jan-1995 #sequence_revision 12-May-1995 #text_change 22-Jun-1999
 C:Accession: S47221
 A:Reference number: S47221
 A:Molecule type: DNA
 A:Residues: 1-452 <MAN1>
 A:Cross-references: EMBL:X81098
 R:Manen, J.F.; Natall, A.; Ehrendorfer, F.
 Submitted to the EMBL Data Library, December 1993
 C:Species: Chloroplast Gallium lucidum
 C:Date: 06-Jan-1995 #sequence_revision 12-May-1995 #text_change 22-Jun-1999
 C:Accession: S47231
 A:Reference number: S47231
 A:Molecule type: DNA
 A:Residues: 1-52, 'X', 54-57 <MAN2>
 A:Cross-references: EMBL:X76461; NID:9434869; PIDN:CA53999.1; PID:9434870
 C:Genetics:
 A:Gene: rbcL
 A:Genome: chloroplast
 C:Function:
 A:Description: catalyzes the reaction of carbon dioxide with D-ribulose 1,5-bisphosphate
 C:Superfamily: ribulose-bisphosphate carboxylase large chain
 C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase;
 F:175,334/Active site: Lys (ribulose-bisphosphate-binding) #status predicted
 F:201/Binding site: carbon dioxide (lys) (covalent) (by Rubisco activase) #status predicted
 F:203/Binding site: magnesium (asp) #status predicted

Query Match 1.6%; Score 8; DB 2; Length 452;
 Best Local Similarity 100.0%; Pred. No. 7.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 IFGDSVL 322
 Db 393 IFGDSVL 400

RESULT 12
 S47230
 ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain - Gallium lucidum chloroplast
 C:Species: Chloroplast Gallium lucidum
 C:Date: 06-Jan-1995 #sequence_revision 12-May-1995 #text_change 22-Jun-1999
 C:Accession: S47230; S39280
 R:Manen, J.F.; Natall, A.
 Submitted to the EMBL Data Library, July 1994
 A:Description: Comparison of the evolution of the rbcL sequence and of the atpB-rbcL not
 A:Reference number: S47221
 A:Accession: S47230
 A:Molecule type: DNA
 A:Residues: 1-452 <MAN1>
 A:Cross-references: EMBL:X81100
 R:Manen, J.F.; Natall, A.; Ehrendorfer, F.
 Submitted to the EMBL Data Library, December 1993
 A:Description: Phylogeny of Rubiaceae-Rubia inferred from the sequence of a cpDNA inter
 A:Reference number: S39270
 A:Accession: S39280
 A:Molecule type: DNA
 A:Residues: 1-52, 'X', 54-57 <MAN2>
 A:Cross-references: EMBL:X76468; NID:9434871; PIDN:CA54006.1; PID:9434872
 C:Genetics:
 A:Gene: rbcL
 A:Genome: chloroplast
 C:Function:
 A:Description: catalyzes the reaction of carbon dioxide with D-ribulose 1,5-bisphosphate
 C:Superfamily: ribulose-bisphosphate carboxylase large chain
 C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase;
 F:175,334/Active site: Lys (ribulose-bisphosphate-binding) #status predicted
 F:201/Binding site: carbon dioxide (lys) (covalent) (by Rubisco activase) #status predicted
 F:203/Binding site: magnesium (asp) #status predicted

Query Match 1.6%; Score 8; DB 2; Length 452;
 Best Local Similarity 100.0%; Pred. No. 7.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 IFGDSVL 322
 Db 393 IFGDSVL 400

RESULT 13
 S47231
 ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain - Gallium palustre chlor
 C:Species: Chloroplast Gallium palustre
 C:Date: 06-Jan-1995 #sequence_revision 12-May-1995 #text_change 22-Jun-1999
 C:Accession: S47231; S39283
 R:Manen, J.F.; Natall, A.
 Submitted to the EMBL Data Library, July 1994
 A:Description: Comparison of the evolution of the rbcL sequence and of the atpB-rbcL
 A:Reference number: S47221
 A:Accession: S47231
 A:Molecule type: DNA
 A:Residues: 1-452 <MAN1>
 A:Cross-references: EMBL:X81102
 R:Manen, J.F.; Natall, A.; Ehrendorfer, F.
 Submitted to the EMBL Data Library, December 1993
 A:Description: Phylogeny of Rubiaceae-Rubia inferred from the sequence of a cpDNA in
 A:Reference number: S39270
 A:Accession: S39283
 A:Molecule type: DNA
 A:Residues: 1-52, 'X', 54-57 <MAN2>
 A:Cross-references: EMBL:X76464; NID:9434877; PIDN:CA54002.1; PID:9434878
 C:Genetics:
 A:Gene: rbcL
 A:Genome: chloroplast
 C:Function:
 A:Description: catalyzes the reaction of carbon dioxide with D-ribulose 1,5-bisphosph
 C:Superfamily: ribulose-bisphosphate carboxylase large chain
 C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase
 F:175,334/Active site: Lys (ribulose-bisphosphate-binding) #status predicted
 F:201/Binding site: carbon dioxide (lys) (covalent) (by Rubisco activase) #status pre
 F:203/Binding site: magnesium (asp) #status predicted

Query Match 1.6%; Score 8; DB 2; Length 452;
 Best Local Similarity 100.0%; Pred. No. 7.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 IFGDSVL 322
 Db 393 IFGDSVL 400

RESULT 14
 S47232
 ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain - Gallium palustre chlorop
 C:Species: Chloroplast Gallium palustre
 C:Date: 06-Jan-1995 #sequence_revision 12-May-1995 #text_change 22-Jun-1999
 C:Accession: S47232; S39281
 R:Manen, J.F.; Natall, A.
 Submitted to the EMBL Data Library, July 1994
 A:Description: Comparison of the evolution of the rbcL sequence and of the atpB-rbcL
 A:Reference number: S47221
 A:Accession: S47232
 A:Molecule type: DNA
 A:Residues: 1-452 <MAN1>
 A:Cross-references: EMBL:X81101
 R:Manen, J.F.; Natall, A.; Ehrendorfer, F.
 Submitted to the EMBL Data Library, December 1993
 A:Description: Phylogeny of Rubiaceae-Rubia inferred from the sequence of a cpDNA in
 A:Reference number: S39270
 A:Accession: S39281
 A:Molecule type: DNA
 A:Residues: 1-52, 'X', 54-57 <MAN2>
 A:Cross-references: EMBL:X76466; NID:9434873; PIDN:CA54004.1; PID:9434874
 C:Genetics:
 A:Gene: rbcL
 A:Genome: chloroplast
 C:Function:

A:Description: catalyzes the reaction of carbon dioxide with D-ribulose 1,5-bisphosphate
 C:Superfamily: ribulose-bisphosphate carboxylase large chain
 C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase; c
 F:175,334/Active site: Lys (ribulose-bisphosphate-binding) #status predicted
 F:201/Binding site: carbon dioxide (Lys) (covalent) (by Rubisco activase) #status predic
 F:203/Binding site: magnesium (Asp) #status predicted

Query Match 1.6%; Score 8; DB 2; Length 452;
 Best Local Similarity 100.0%; Pred.No. 7.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 315 IFGDDSVL 322
 DB 393 IFGDDSVL 400

RESULT 15

S47233
 ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain - Gallium scabrum chloroplast
 C:Species: chloroplast Gallium scabrum
 C:Date: 06-Jan-1995 #sequence_revision 12-May-1995 #text_change 22-Jun-1999
 C:Accession: S47233; S39284
 R:Manen, J.F.; Natali, A.
 Submitted to the EMBL Data Library, July 1994
 A:Description: Comparison of the evolution of the rbcl sequence and of the atpB-rbcl non
 A:Reference number: S47221

A:Accession: S47233

A:Molecule type: DNA

A:Residues: 1452 <MAN1>

A:Cross-references: EMBL:X81105

R:Manen, J.F.; Natali, A.; Ehendorfer, F.
 Submitted to the EMBL Data Library, December 1993

A:Description: Phylogeny of Rubiaceae-Rubia inferred from the sequence of a cpDNA inter

A:Reference number: S39270

A:Accession: S39284

A:Molecule type: DNA

A:Residues: 1-52, 'X', 54-57 <MAN2>

A:Cross-references: EMBL:X76462; NID:9434879; PIDN:CA54000.1; PID:9434880

C:Genetics:

A:Gene: rbcl

A:Genome: chloroplast

C:Function:

A:Description: catalyzes the reaction of carbon dioxide with D-ribulose 1,5-bisphosphate

C:Superfamily: ribulose-bisphosphate carboxylase large chain

C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase; c

F:175,334/Active site: Lys (ribulose-bisphosphate-binding) #status predicted

F:201/Binding site: carbon dioxide (Lys) (covalent) (by Rubisco activase) #status predic

F:203/Binding site: magnesium (Asp) #status predicted

OY 315 IFGDDSVL 322
 DB 393 IFGDDSVL 400

Search completed: November 29, 2002, 12:46:52
 Job time : 24 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2002, 12:37:40 ; Search time 17 Seconds

(without alignments)
1212.573 Million cell updates/sec

Title: US-09-868-760-7

Perfect score: 497
Sequence: 1 DGEQNDGQNKDHHDDHD.....MOGHEVESERVACCVIGRA 497

Scoring table: OLIGO
Gapop 60.0, Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size: 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Listing first 45 summaries

Database: SwissProt.40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.8	471	1 RBL2_HYDMR	Q59460 hydrogenvi
2	8	1.6	25	1 P706_BOVIN	P82931 bos taurus
3	8	1.6	125	1 P706_MOUSE	P58064 mus musculu
4	8	1.6	149	1 EGGS_SCHMA	P08016 schistosoma
5	8	1.6	394	1 RBL_ALIPV	P34767 allisma plan
6	8	1.6	394	1 RBL_BRALO	Q05795 barclaya lo
7	8	1.6	394	1 RBL_BRASC	Q05796 brasenia sc
8	8	1.6	394	1 RBL_CABCA	Q05797 cabomba car
9	8	1.6	394	1 RBL_CERDE	Q05798 ceratophyll
10	8	1.6	394	1 RBL_EURPE	Q05799 eurypale fer
11	8	1.6	394	1 RBL_NELIU	Q05800 nelumbo lut
12	8	1.6	394	1 RBL_NYMPA	Q05801 nymphea vari
13	8	1.6	394	1 RBL_NYMPD	Q05802 nymphea od
14	8	1.6	406	1 RBL_VICCZ	Q05803 victoria cr
15	8	1.6	410	1 RBL_CREBI	P48729 clepidomane
16	8	1.6	413	1 RBL_GLBJA	P43223 adiantum pe
17	8	1.6	414	1 RBL_BUEOR	P43226 blechnum pe
18	8	1.6	414	1 RBL_CONYA	Q36510 onychium ja
19	8	1.6	415	1 RBL_CITBA	P43228 clobutium ba
20	8	1.6	416	1 RBL_ARTBA	P43228 arthroperti
21	8	1.6	416	1 RBL_SPIRK	P36488 spigelia ma
22	8	1.6	420	1 RBL_ANEME	Q31674 anemia mexi
23	8	1.6	421	1 RBL_AEGRI	P25113 aegilops cr
24	8	1.6	421	1 RBL_AEGRI	P25114 aegilops ta
25	8	1.6	422	1 RBL_AEGRI	P25114 aegilops ta
26	8	1.6	422	1 RBL_AEGRI	P25114 aegilops ta
27	8	1.6	422	1 RBL_AEGRI	P25114 aegilops ta
28	8	1.6	422	1 RBL_AEGRI	P25114 aegilops ta
29	8	1.6	422	1 RBL_AEGRI	P25114 aegilops ta
30	8	1.6	422	1 RBL_AEGRI	P25114 aegilops ta
31	8	1.6	422	1 RBL_AEGRI	P25114 aegilops ta
32	8	1.6	422	1 RBL_AEGRI	P25114 aegilops ta
33	8	1.6	422	1 RBL_AEGRI	P25114 aegilops ta

ALIGNMENTS

RESULT 1
ID RBL2_HYDMR STANDARD: PRT: 471 AA.
AC Q59460;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE Ribulose biphosphate carboxylase large chain 2 (EC 4.1.1.39)
DE (Rubisco large subunit).
GN CBL-2.
OS Hydrogenovibrio marinus.
OC Bacteria; Proteobacteria; gamma subdivision; Piscirickettsia group;
OC Hydrogenovibrio.
OX NCBI_TaxID=28885;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MH-110;
RA Yeguchi T., Chung S., Yasuo I., Tohru K.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE.
CC -1- CATALYTIC ACTIVITY: D-Ribulose 1,5-bisphosphate + CO(2) = 2 3-
CC phospho-D-glycerate.
CC -1- CATALYTIC ACTIVITY: D-Ribulose 1,5-bisphosphate + O(2) =
CC 3-phospho-D-glycerate + 2-phosphoglycolate.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D43622; BAA07731.1;
DR HSSP: P09657; IERN.
DR InterPro: IPR000685; Rubisco_large.
DR Pfam: PF00016; Rubisco_large; 1.
DR Pfam: PF02788; Rubisco_large; 1.
DR PROSITE: PS00157; RUBISCO_LARGE; FALSE_NEG.
KW Carbon dioxide fixation; Photorespiration; Lyase; Oxidoreductase;
KW Monooxygenase; Calvin cycle; Multigene family.
FT ACT_SITE 194
FT BINDING OF CO(2) ACTIVATES THE ENZYME
FT (BY SIMILARITY).
SQ SEQUENCE 471 AA: 52005 MW; BAD462E46451AD50 CRC64;
Query Match 1.8%; Score 9; DB 1; Length 471;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 NIFGDDSVL 322
 |||||||
 DB 385 NIFGDDSVL 393

RESULT 2

RT06_BOVIN STANDARD: PRT: 25 AA.
 AC P82931;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Mitochondrial 28S ribosomal protein S6 (MRP-S6) (Fragments).
 GN MRP56 OR RRP56.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Kuntunalia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 ON NCBI_TaxID=9913;
 RX [1]
 RP SEQUENCE:
 RC TISSUE=Liver;
 RA MEDLINE=21276436; PubMed=11279123;
 RA Koc E.C.; Burkhardt W.; Blackburn K.; Moseley A.; Spremulli L.L.;
 RT "The small subunit of the mammalian mitochondrial ribosome:
 RT Identification of the full complement of ribosomal proteins present."
 RU J. Biol. Chem. 276:19363-19374(2001).
 CC -1- SUBCELLULAR LOCATION: Mitochondrial.
 CC -1- SIMILARITY: BELONGS TO THE S6P FAMILY OF RIBOSOMAL PROTEINS.
 DR InterPro: IPR000529; Ribosomal_S6.
 DR Pfam: PFO1250; Ribosomal_S6; PARTIAL.
 DR PROSITE: PS01048; RIBOSOMAL_S6; PARTIAL.
 KW Ribosomal protein; Mitochondrion.
 FT NON_TER 1 1
 FT NON_CONS 12 13
 FT NON_TER 25 25
 SQ SEQUENCE 25 AA; 2710 MW; 57C89F522C889970 CRC64;

Query Match 1.6%; Score 8; DB 1; Length 25;
 Best Local Similarity 100.0%; Pred. No. 0.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 187 RPERAAL 194
 |||||||
 DB 4 RPERAAL 11

RESULT 3

RT06_MOUSE STANDARD: PRT: 125 AA.
 AC P58064;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Mitochondrial 28S ribosomal protein S6 (MRP-S6).
 GN MRP56 OR RRP56.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA SUZUKI T.; Terasaki M.; Takemoto C.; Hanada T.; Wada A.; Ueda T.;
 RA Watanabe K.;
 RT "Mammalian mitochondrial ribosome: structural and functional
 RT compensation for deficit of RNA with protein components."
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Mitochondrial.
 CC -1- SIMILARITY: BELONGS TO THE S6P FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AB049943; BAB40996.1; ALT_INIT.
 DR InterPro: IPR000529; Ribosomal_S6.
 DR TIGRFAMs: TIGR00166; S6; 1.
 DR PROSITE: PS01048; RIBOSOMAL_S6; FALSE_NEG.
 KW Ribosomal protein; Mitochondrion.
 SQ SEQUENCE 125 AA; 14308 MW; 8140DB0BFPDP9025 CRC64;

Query Match 1.6%; Score 8; DB 1; Length 125;
 Best Local Similarity 100.0%; Pred. No. 1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 187 RPERAAL 194
 |||||||
 DB 15 RPERAAL 22

RESULT 4

EGGS_SCHEMA STANDARD: PRT: 149 AA.
 AC P08016;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE Putative eggshell protein (fragment).
 OS Schistosoma mansoni (Blood fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeiidae;
 OC Schistosomatidae; Schistosomatidae; Schistosoma.
 ON NCBI_TaxID=6183;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=87115624; PubMed=3100949;
 RA Johnson K.S.; Taylor D.W.; Cordingley J.S.;
 RT "Possible eggshell protein gene from Schistosoma mansoni."
 RL Mol. Biochem. Parasitol. 22:89-100(1987).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: M15371; AAA29877.1; -.
 DR PIR: A54530; A54530.
 KW Eggshell; Repeat.
 FT NON_TER 1 1
 FT DOMAIN 109 145 HIS-RICH.
 SQ SEQUENCE 149 AA; 18839 MW; F5399D570845274A CRC64;

Query Match 1.6%; Score 8; DB 1; Length 149;
 Best Local Similarity 100.0%; Pred. No. 1.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 DHHDHHD 20
 |||||||
 DB 118 DHHDHHD 125

RESULT 5

RBL_ALIPL STANDARD: PRT: 394 AA.
 AC P34767;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

RA Les D.H., Garvin D.K., Wimpsee C.F.;
RT "Molecular evolutionary history of ancient aquatic angiosperms."
RL Proc. Natl. Acad. Sci. U.S.A. 88:10119-10123(1991).
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) -
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
CC D-RIBULOSE, 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) - 2 3-
CC phospho-D-glycerate.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) -
CC 3-phospho-D-glycerate + 2-phosphoglycolate.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M77031; AAA84085.1; -
DR InterPro: IPR000685; Rubisco_large.
DR Pfam: PF00016; Rubisco_large; 1.
DR Pfam: PF02788; Rubisco_large; N: 1.
DR PROSITE: PS00157; RUBISCO_LARGE; 1.
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.
FT NON_TER 1 1
FT ACT_SITE 192 192 BINDING OF CO(2) ACTIVATES THE ENZYME.
FT NON_TER 394 394
SQ SEQUENCE 394 AA: 43778 MW: 36DB85FE4408082F CRC64;

Query Match 1.6%; Score 8; DB 1; Length 394;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 315 IFGDSYVL 322
DB 384 IFGDSYVL 391

RESULT 8
RBL_CARCA STANDARD; PRT; 394 AA.
ID RBL_CARCA
AC Q05797; 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO
DE large subunit) (Fragment).
GN RBL.
OS Cabomba caroliniana.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Cabombaceae; Cabomba.
OX NCBI_TaxID=4426;
RN [1]
RP SEQUENCE FROM N.A.
RA Les D.H., Garvin D.K., Wimpsee C.F.;
RX MEDLINE=92052222; PubMed=1946432;
RT "Molecular evolutionary history of ancient aquatic angiosperms."
RL Proc. Natl. Acad. Sci. U.S.A. 88:10119-10123(1991).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE.
CC -----

CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) - 2 3-
CC phospho-D-glycerate.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) -
CC -1- 3-phospho-D-glycerate + 2-phosphoglycolate.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M77027; AAA84098.1; -
DR InterPro: IPR000685; Rubisco_large.
DR Pfam: PF00016; Rubisco_large; 1.
DR Pfam: PF02788; Rubisco_large; N: 1.
DR PROSITE: PS00157; RUBISCO_LARGE; 1.
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.
FT NON_TER 1 1
FT ACT_SITE 192 192 BINDING OF CO(2) ACTIVATES THE ENZYME.
FT NON_TER 394 394
SQ SEQUENCE 394 AA: 43810 MW: B747E1471771EBB3 CRC64;

Query Match 1.6%; Score 8; DB 1; Length 394;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 315 IFGDSYVL 322
DB 384 IFGDSYVL 391

RESULT 9
RBL_CERDE STANDARD; PRT; 394 AA.
ID RBL_CERDE
AC Q05798; 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO
DE large subunit) (Fragment).
GN RBL.
OS Ceratophyllum demersum.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Ceratophyllales; Ceratophyllaceae;
OC Ceratophyllum.
OX NCBI_TaxID=4428;
RN [1]
RP SEQUENCE FROM N.A.
RA Les D.H., Garvin D.K., Wimpsee C.F.;
RX MEDLINE=92052222; PubMed=1946432;
RT "Molecular evolutionary history of ancient aquatic angiosperms."
RL Proc. Natl. Acad. Sci. U.S.A. 88:10119-10123(1991).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) - 2 3-
CC phospho-D-glycerate.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) -
CC -1- 3-phospho-D-glycerate + 2-phosphoglycolate.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M77030; AAA84104.1; -
DR InterPro: IPR000685; Rubisco_large.
DR Pfam: PF00016; Rubisco_large; 1.
DR Pfam: PF02788; Rubisco_large; N; 1.
DR PROSITE: PS00157; RUBISCO_LARGE; 1.
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.
FT NON_TER 1 192
FT ACT_SITE 192 192 BINDING OF CO(2) ACTIVATES THE ENZYME.
FT NON_TER 394 394
SQ SEQUENCE 394 AA; 43948 MW; 70078C4D989194D5 CRC64;

Query Match 1.6%; Score 8; DB 1; Length 394;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 315 IFGDDSVL 322
Db 384 IFGDDSVL 391
| | | | | | | |
RBL_EURPE STANDARD; PRT; 394 AA.
ID RBL_EURPE
AC 005799;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (Rubisco
DE large subunit) (Fragment).
GN RBCL.
OS Euryale ferox.
OC Chloroplast.
CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Magnoliophyta; Nymphaeaceae; Euryale.
CX NCBI_TaxID:4414;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE:92052222; PubMed:1946432;
RA Les D.H., Garvin D.K., Wimpsee C.F.;
RT "Molecular evolutionary history of ancient aquatic angiosperms.";
Proc. Natl. Acad. Sci. U.S.A. 88:10119-10123(1991).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) = 2 3-
CC phospho-D-glycerate.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) =
CC 3-phospho-D-glycerate + 2-phosphoglycolate.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
CC -1- SUBCELLULAR LOCATION: Chloroplast.
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M77035; AAA84249.1; -
DR InterPro: IPR000685; Rubisco_large.
DR Pfam: PF00016; Rubisco_large; 1.
DR Pfam: PF02788; Rubisco_large; N; 1.
DR PROSITE: PS00157; RUBISCO_LARGE; 1.
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.
FT NON_TER 1 192
FT ACT_SITE 192 192 BINDING OF CO(2) ACTIVATES THE ENZYME.
FT NON_TER 394 394

DR InterPro: IPR000685; Rubisco_large.
DR Pfam: PF00016; Rubisco_large; 1.
DR Pfam: PF02788; Rubisco_large; N; 1.
DR PROSITE: PS00157; RUBISCO_LARGE; 1.
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.
FT NON_TER 1 192
FT ACT_SITE 192 192 BINDING OF CO(2) ACTIVATES THE ENZYME.
FT NON_TER 394 394
SQ SEQUENCE 394 AA; 43862 MW; 3D9FA4821BF8A143D CRC64;

Query Match 1.6%; Score 8; DB 1; Length 394;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 315 IFGDDSVL 322
Db 384 IFGDDSVL 391
| | | | | | | |
RBL_NELTU STANDARD; PRT; 394 AA.
ID RBL_NELTU
AC 005800;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (Rubisco
DE large subunit) (Fragment).
GN RBCL.
OS Nelumbo lutea (American lotus).
OC Chloroplast.
CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Magnoliophyta; eudicotyledons; Nelumbonaceae; Nelumbo.
CX NCBI_TaxID:4431;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE:92052222; PubMed:1946432;
RA Les D.H., Garvin D.K., Wimpsee C.F.;
RT "Molecular evolutionary history of ancient aquatic angiosperms.";
Proc. Natl. Acad. Sci. U.S.A. 88:10119-10123(1991).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) = 2 3-
CC phospho-D-glycerate.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) =
CC 3-phospho-D-glycerate + 2-phosphoglycolate.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
CC -1- SUBCELLULAR LOCATION: Chloroplast.
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M77032; AAA84499.1; -
DR InterPro: IPR000685; Rubisco_large.
DR Pfam: PF00016; Rubisco_large; 1.
DR Pfam: PF02788; Rubisco_large; N; 1.
DR PROSITE: PS00157; RUBISCO_LARGE; 1.
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.
FT NON_TER 1 192
FT ACT_SITE 192 192 BINDING OF CO(2) ACTIVATES THE ENZYME.
FT NON_TER 394 394

SO SEQUENCE 394 AA; 43917 MW; 93295743D36D2E4E CRC64;

Query Match 1.6%; Score 8; DB 1; Length 394;
 Best Local Similarity 100.0%; Pred. No. 2.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 315 IFGDDSVL 322
 |||||||
 DB 384 IFGDDSVL 391

RESULT 12

ID	RBL_NUPVA	STANDARD	PRT	394 AA
AC	005801;			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO large subunit) (Fragment).			
GN	RBCl.			
OS	Nuphar variegata (Yellow pond lily).			
OC	Chloroplast.			
CC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
CC	Spermatophyta; Magnoliophyta; Nymphaeaceae; Nuphar.			
CC	NCBI_TaxID=4416;			
CC	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92052222; PubMed=1946432;			
RA	Les D.H., Garvin D.K., Winsee C.F.;			
RT	"Molecular evolutionary history of ancient aquatic angiosperms.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 88:10119-10123(1991).			
CC	-1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME ACTIVE SITE.			
CC	-1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) -> 2 3-phospho-D-glycerate.			
CC	-1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) -> 3-phospho-D-glycerate + 2-phosphoglycolate.			
CC	-1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.			
CC	-1- SUBCELLULAR LOCATION: Chloroplast.			
CC	-1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: M77029; AAA84527.1; -			
DR	InterPro: IPR000685; RUBISCO_large.			
DR	Pfam: PF00016; RUBISCO_large; 1.			
DR	Pfam: PF02788; RUBISCO_large; N: 1.			
DR	PROSITE: PS00157; RUBISCO_LARGE; 1.			
KW	Photosynthesis; Carbon dioxide fixation; Photorespiration;			
KW	Lysase; Oxidoreductase; Monooxygenase; Chloroplast.			
FT	NON_TER 1			
FT	ACT_SITE 192 192			
FT	NON_TER 394 394			
SO	SEQUENCE 394 AA; 43816 MW; E400510ADB98C803 CRC64;			

Query Match 1.6%; Score 8; DB 1; Length 394;
 Best Local Similarity 100.0%; Pred. No. 2.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 315 IFGDDSVL 322
 |||||||
 DB 384 IFGDDSVL 391

RESULT 13

ID	RBL_NYMOD	STANDARD	PRT	394 AA
AC	005802;			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO large subunit) (Fragment).			
GN	RBCl.			
OS	Nymphaea odorata (White water lily).			
OC	Chloroplast.			
CC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
CC	Spermatophyta; Magnoliophyta; Nymphaeaceae; Nymphaea.			
CC	NCBI_TaxID=4419;			
CC	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92052222; PubMed=1946432;			
RA	Les D.H., Garvin D.K., Winsee C.F.;			
RT	"Molecular evolutionary history of ancient aquatic angiosperms.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 88:10119-10123(1991).			
CC	-1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME ACTIVE SITE.			
CC	-1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) -> 2 3-phospho-D-glycerate.			
CC	-1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) -> 3-phospho-D-glycerate + 2-phosphoglycolate.			
CC	-1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.			
CC	-1- SUBCELLULAR LOCATION: Chloroplast.			
CC	-1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: M77034; AAA84529.1; -			
DR	InterPro: IPR000685; RUBISCO_large.			
DR	Pfam: PF00016; RUBISCO_large; 1.			
DR	Pfam: PF02788; RUBISCO_large; N: 1.			
DR	PROSITE: PS00157; RUBISCO_LARGE; 1.			
KW	Photosynthesis; Carbon dioxide fixation; Photorespiration;			
KW	Lysase; Oxidoreductase; Monooxygenase; Chloroplast.			
FT	NON_TER 1			
FT	ACT_SITE 192 192			
FT	NON_TER 394 394			
SO	SEQUENCE 394 AA; 43832 MW; 31E3C25FBBD0A7EF CRC64;			

Query Match 1.6%; Score 8; DB 1; Length 394;
 Best Local Similarity 100.0%; Pred. No. 2.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 315 IFGDDSVL 322
 |||||||
 DB 384 IFGDDSVL 391

RESULT 14

ID	RBL_VIC2	STANDARD	PRT	394 AA
AC	005803;			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			


```

DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO
DE large subunit) (Fragment).
GN RBCL.
OS Victoria cruziana (Santa Cruz water lily).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Nymphaeaceae; Victoria.
ON NCBI_TaxID=4421;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92052222; PubMed=1946432;
RA Les D.H., Garin D.K., Wimpsee C.F.;
RT "Molecular evolutionary history of ancient aquatic angiosperms.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:10119-10123(1991).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) = 2 3-
CC phospho-D-glycerate.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) =
CC 3-phospho-D-glycerate + 2-phosphoglycolate.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
CC -1- SUBCELLULAR LOCATION: Chloroplast.
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M7036; AA84716.1;
DR InterPro: IPR000685; RUBISCO_large.
DR Pfam: PF00016; RUBISCO_large; 1.
DR Pfam: PF02788; RUBISCO_large; 1.
DR PROSITE: PS00157; RUBISCO_LARGE; 1.
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.
FT NON_TER 1 1
FT ACT_SITE 192 192 BINDING OF CO(2) ACTIVATES THE ENZYME.
FT NON_TER 394 394
SQ SEQUENCE 394 AA; 43848 MW; 13EF14825FEAE13D CRC64;

Query Match 1.6%; Score 8; DB 1; Length 394;
Best Local Similarity 100.0%; Pred. No. 2; 9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 315 IFGDDSVL 322
DB 384 IFGDDSVL 391

```

```

RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;
RX MEDLINE=94261660; PubMed=8202555;
RA Hasebe M., Omori T., Nakazawa M., Sano T., Kato M., Iwatsuki K.;
RT "rbcL gene sequences provide evidence for the evolutionary lineages
RT of leptosporangiate ferns.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:5730-5734(1994).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) = 2 3-
CC phospho-D-glycerate.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) =
CC 3-phospho-D-glycerate + 2-phosphoglycolate.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
CC -1- SUBCELLULAR LOCATION: Chloroplast.
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U05613; AAA1897.1;
DR InterPro: IPR000685; RUBISCO_large.
DR Pfam: PF00016; RUBISCO_large; 1.
DR Pfam: PF02788; RUBISCO_large; 1.
DR PROSITE: PS00157; RUBISCO_LARGE; 1.
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.
FT NON_TER 1 1
FT ACT_SITE 179 179 BINDING OF CO(2) ACTIVATES THE ENZYME.
FT NON_TER 406 406
SQ SEQUENCE 406 AA; 44888 MW; 47137F9AED3F7950 CRC64;

Query Match 1.6%; Score 8; DB 1; Length 406;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 315 IFGDDSVL 322
DB 371 IFGDDSVL 378

Search completed: November 29, 2002, 12:45:18
Job time : 20 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2002, 12:38:35 : Search time 32 Seconds

(without alignments)
3200.169 Million cell updates/sec

Title: US-09-868-760-7

Perfection score: 497
Sequence: 1 DGECCNDGQKNDHDDHDD.....MOGHEVESERVACVIGRA 497

Scoring table:
Gapop 60.0, Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size: 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: SPTREMBL_21:*

1: sp-archaea:*

2: sp-bacteria:*

3: sp-fungi:*

4: sp-human:*

5: sp-invertebrate:*

6: sp-mammal:*

7: sp-mhc:*

8: sp-organelle:*

9: sp-phage:*

10: sp-plant:*

11: sp-proteob:*

12: sp-virus:*

13: sp-vertebrate:*

14: sp-unclassified:*

15: sp-virus:*

16: sp-bacteriap:*

17: sp-archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	434	87.3	517	5	Q9BKB9
2	9	1.8	20	5	P83148
3	9	1.8	79	17	Q9HPN6
4	9	1.8	473	2	O85040
5	9	1.8	473	2	O85040
6	8	1.6	83	2	O86380
7	8	1.6	83	2	O86380
8	8	1.6	83	2	O86380
9	8	1.6	83	2	O86380
10	8	1.6	83	2	O86380
11	8	1.6	83	2	O86380
12	8	1.6	83	2	O86380
13	8	1.6	83	2	O86380
14	8	1.6	83	2	O86380
15	8	1.6	83	2	O86380
16	8	1.6	83	2	O86380

17	8	1.6	83	2	O939W0
18	8	1.6	83	2	O939V5
19	8	1.6	83	2	O939V4
20	8	1.6	83	2	O939V3
21	8	1.6	83	2	O934F1
22	8	1.6	83	2	O934F0
23	8	1.6	83	2	O934F9
24	8	1.6	83	2	O934F8
25	8	1.6	83	2	O934F7
26	8	1.6	83	2	O934F6
27	8	1.6	83	2	O934F5
28	8	1.6	83	2	O934F4
29	8	1.6	83	2	O934F3
30	8	1.6	83	2	O934F2
31	8	1.6	83	2	O934F1
32	8	1.6	83	2	O934F0
33	8	1.6	83	2	O934F9
34	8	1.6	83	2	O934F8
35	8	1.6	83	2	O934F7
36	8	1.6	83	2	O934F6
37	8	1.6	83	2	O934F5
38	8	1.6	83	2	O934F4
39	8	1.6	83	2	O934F3
40	8	1.6	83	2	O934F2
41	8	1.6	83	2	O934F1
42	8	1.6	83	2	O934F0
43	8	1.6	83	2	O934F9
44	8	1.6	83	2	O934F8
45	8	1.6	83	2	O934F7

ALIGNMENTS

RESULT 1

ID	Q9BKB9	PRELIMINARY:	PRT:	517 AA.
AC	O9BKB9			
DI	01-JUN-2001 (TREMBLER, 17, Created)			
DI	01-JUN-2001 (TREMBLER, 17, Last sequence update)			
DI	01-JUN-2002 (TREMBLER, 21, Last annotation update)			
DE	Perna precursor.			
OS	Perna canaliculus (greenshell mussel).			
OC	Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;			
OC	Mytiloidea; Mytilidae; Perna.			
OX	NCBI_TaxID=38949;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21186417; PubMed=11290459;			
RA	Scotti P.D., Dearing S.C., Greenwood D.R., Newcomb R.D.;			
RT	"Perna: a novel self-aggregating haemolymph protein from the New			
RT	Zealand green-lipped mussel Perna canaliculus (bivalvia: mytilidae)."			
RL	Comp. Biochem. Physiol. B, Biochem. Mol. Biol. 128:767-779(2001).			
CC	- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE			
CC	CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).			
CC	- CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) = O(2) + H(2)O(2).			
CC	- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.			
DR	EMBL: AF273766; AAK20952.1; -			
DR	HSSP: P00445; 1JCV.			
DR	InterPro: IPR001424; SOD_CU_ZN.			
DR	Pfam: PF00080; sodbu.3			
DR	PRINTS: PR00068; CUZNDISMUTASE.			
DR	ProDom: PD000469; SOD_CU_ZN.3			
KW	Copper; Oxidoreductase; Signal; 2inc.			
FT	SIGNAL 1 20			
FT	CHAIN 21 517			
FT	POTENTIAL.			
FT	PERIN.			
SQ	SEQUENCE 517 AA: 57222 MW; 8788F8FE85501E CRC64;			
Query Match	87.3%; Score 434; DB 5; Length 517;			
Best Local Similarity	100.0%; Pred. No. 0;			
Matches 434; Conservative	0; Mismatches 0; Indels 0; Gaps 0;			
OY	64 GHGAVYLIELHLVGFNTSEDDHDDHGHGLHMLGDMASGDSIGELYNAPPEKADPDGLG 123			

DB 84 GGGAGVLEHLVGNPTSEDDHHHGLHMLHMDMSAGDSIGELYNAHPKRNHPDGLG 143
OY 124 DLVDDRGVYVNEHYAAMDIDGTAPNTEALIGHSMTLLQSSHTDADTPASRIACVYGH 183
DB 144 DLVDDRGVYVNEHYAAMDIDGTAPNTEALIGHSMTLLQSSHTDADTPASRIACVYGH 203
OY 184 GKARPEETAALHNELEDKTEHYAHCDVRSNTHOPKALHHNHGHTIDPKOYGYDLEVS 243
DB 204 GKARPEETAALHNELEDKTEHYAHCDVRSNTHOPKALHHNHGHTIDPKOYGYDLEVS 263
OY 244 HLEGFNVSDDHKDHLHDVQIYANGDLTSGCDNLGAKYDPHEDEHSELGLDIDHDDHGV 303
DB 264 HLEGFNVSDDHKDHLHDVQIYANGDLTSGCDNLGAKYDPHEDEHSELGLDIDHDDHGV 323
OY 304 VNESHRYSMINIFEGDSDVLGRSIAIHQRDLHLSAKIACVYGRGOSHPETVHRAKCYVR 363
DB 324 VNESHRYSMINIFEGDSDVLGRSIAIHQRDLHLSAKIACVYGRGOSHPETVHRAKCYVR 383
OY 364 PNRESTGLHHVSGSTFEQTPGSGTHMTADLKGFVSEDLSSHFGVQLHEKMDMSHGC 423
DB 384 PNRESTGLHHVSGSTFEQTPGSGTHMTADLKGFVSEDLSSHFGVQLHEKMDMSHGC 443
OY 424 HSLGRMTHGHDDAHPKRPDGLDVIDSHGIVHSTRFEDHLNVEDLNARSLVTMGGCHE 483
DB 444 HSLGRMTHGHDDAHPKRPDGLDVIDSHGIVHSTRFEDHLNVEDLNARSLVTMGGCHE 503
OY 484 VESERVACVYIGRA 497
DB 504 VESERVACVYIGRA 517

RESULT 2

P83148 PRELIMINARY; PRT; 20 AA.
AC P83148;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE Major extracellular fluid protein (EP protein) (Fragment).
OS Mytilus edulis (Blue mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;
OC Mytiloidae; Mytilidae; Mytilus.
NCBI_TaxID=6550.

RP SEQUENCE, SUBUNIT, GLYCOSYLATION, AND MASS SPECTROMETRY.
RC TISSUE-EXTRAPALLIAL FLUID.
RX Harten S.J., Laue T.M., Chasteen N.D.;
MEDLINE=21264951; PubMed=11084027;
RT "Purification and characterization of a novel calcium-binding protein
from the extrapallial fluid of the mussel, Mytilus edulis."
RL J. Biol. Chem. 276:4461-4468(2001).
CC -1- FUNCTION: APPEARS TO BE A BUILDING BLOCK OF THE SOLUBLE ORGANIC
MATRIX OF THE SHELL. THE PROTEIN BINDS CALCIUM.
CC -1- SUBUNIT: HOMODIMER.
CC -1- PTM: GLYCOSYLATED.
CC -1- MASS SPECTROMETRY: MW=28340; METHOD=MALDI.
CC -1- MISCELLANEOUS: HAS AN ISOELECTRIC POINT OF 4.43.
KM Calcium-binding; Glycoprotein.
FT NON_TER 20
SQ SEQUENCE 20 AA; 2351 MW; CBB0C9EDDE6F7451 CRC64;

Query Match 1.8%; Score 9; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.085;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 3

O9HPN6
DB 4 DHHDDHND 12

ID O9HPN6 PRELIMINARY; PRT; 79 AA.
AC O9HPN6;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Vn91546h.
GN Vn91546h.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahlras G.G., Bergquist B., Pan M.,
RA Shukla H.D., Laake S.R., Bailga N.S., Thorsson V., Sirogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Weli R., Goo Y.A.,
RA Leitbauer B., Keller K., Cruz R., Danon M.J., Hough D.W.,
RA Leddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenberger T.A., Beck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omar A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarina S.,
RA proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AF005068; AF019831.1; -;
DR InterPro; IPR004045; GST_Nterm.
KW Complete proteome.
SQ SEQUENCE 79 AA; 8511 MW; CAC5748CA97B4BE0 CRC64;

Query Match 1.8%; Score 9; DB 17; Length 79;
Best Local Similarity 100.0%; Pred. No. 0.28;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 125 LVDDRGVY 133
DB 55 LVDDRGVY 63

RESULT 4

O85040 PRELIMINARY; PRT; 473 AA.
ID O85040;
AC O85040;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DE 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Ribulose biphosphate carboxylase large chain (Ec 4.1.1.39) (Rubisco
large subunit).
GN Rbcl OR CBRL.
OS Thiobacillus neapolitanus.
OC Bacteria; Proteobacteria; gamma subdivision; Chromatiales;
OC Halothiobacillus.
NCBI_TaxID=927;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 23641;
RX MEDLINE=98361901; PubMed=9696760;
RA Baker S.H., Jin S., Aldrich H.C., Howard G.T., Shively J.M.;
RT Insertion mutation of the form I cbcl gene encoding ribulose
biphosphate carboxylase/oxygenase (Rubisco) in Thiobacillus
neapolitanus results in expression of form II Rubisco, loss of
RT neapolitanus results in expression of form II Rubisco, loss of
RT carboxysomes, and an increased CO2 requirement for growth.
RL J. Bacteriol. 180:4133-4139(1998).

CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-
RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
ACTIVE SITE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) - 3-
CC -1- PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) - 2 3-
CC -1- PHOSPHO-D-GLYCERATE.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).

```

CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
DR EMBL: AF038430; AAC32549.1; -.
DR HSSP: P09657; IBXN.
DR InterPro: IPR000685; Rubisco_large.
DR Pfam: PF00016; Rubisco_large.1.
DR PROSITE: PS00157; RUBISCO_LARGE; 1.
KW Carbon dioxide fixation; Lyase; Monooxygenase; Oxidoreductase;
KW Photorespiration.
SQ SEQUENCE 473 AA; 52635 MW; B84D2DE46CAF7D8 CRC64;

Query Match 1.8%; Score 9; DB 2; Length 473;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 NIFGDDSVL 322
Db 385 NIFGDDSVL 393

RESULT 5
O92B35 PRELIMINARY; PRT; 473 AA.
AC O92B35;
DT 01-MAY-1999 (TREMBLrel. 10; Created)
DT 01-MAY-1999 (TREMBLrel. 10; Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20; Last annotation update)
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (Rubisco
DE large subunit).
GN RCL.
OS Pseudomonas carboxydoflava.
OC Bacteria; Proteobacteria; beta subdivision; Comamonadaceae;
OC Hydrogenophaga.
OX NCBI_TaxID=47421;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1083;
RA Kim I.M.; Lee S.N.;
RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-
CC RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-
CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-
CC PHOSPHO-D-GLYCERATE.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
DR EMBL: U55037; AAD1031.1; -.
DR InterPro: IPR000685; Rubisco_large.
DR Pfam: PF00016; Rubisco_large.1.
DR PROSITE: PS00157; RUBISCO_LARGE; 1.
KW Carbon dioxide fixation; Lyase; Monooxygenase; Oxidoreductase;
KW Photorespiration.
SQ SEQUENCE 473 AA; 52692 MW; 87F82B96F8A35C59 CRC64;

Query Match 1.8%; Score 9; DB 2; Length 473;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 NIFGDDSVL 322
Db 385 NIFGDDSVL 393

RESULT 6
O86380 PRELIMINARY; PRT; 83 AA.
AC O86380;

```

```

DT 01-NOV-1998 (TREMBLrel. 08; Created)
DT 01-NOV-1998 (TREMBLrel. 08; Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20; Last annotation update)
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (Rubisco
DE large subunit) (Fragment).
GN RCL.
OS Nostoc commune.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=1178;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98317287; PubMed=9642201;
RA Rudi K.; Skulberg O.M.; Jakobsen K.S.;
RT "Evolution of cyanobacteria by exchange of genetic material among
RT phylogenetically related strains."
RL J. Bacteriol. 180:3453-3461(1998).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-
CC RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-
CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-
CC PHOSPHO-D-GLYCERATE.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
DR EMBL: Z94892; CAB08231.1; -.
DR InterPro: IPR000685; Rubisco_large.
DR Pfam: PF00016; Rubisco_large.1.
KW Carbon dioxide fixation; Lyase; Monooxygenase; Oxidoreductase;
KW Photorespiration.
FT NONTER 1
SQ SEQUENCE 83 AA; 9013 MW; 00F555C851F25D CRC64;

Query Match 1.6%; Score 8; DB 2; Length 83;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 315 IFGDDSVL 322
Db 1 IFGDDSVL 8

RESULT 7
O93A05 PRELIMINARY; PRT; 83 AA.
AC O93A05;
DT 01-DEC-2001 (TREMBLrel. 19; Created)
DT 01-DEC-2001 (TREMBLrel. 19; Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20; Last annotation update)
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (Rubisco
DE large subunit) (Fragment).
GN RCL.
OS Anabaena cf. lemmermannii 262.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
OX NCBI_TaxID=135961;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=262;
RA Guiger M.; Lyra C.; Henriksen P.; Coute A.; Humbert J.F.; Sivonen K.;
RT "Morphological and molecular comparison of the cyanobacterial genera
RT Anabaena and Aphanizomenon."
RL Int. J. Syst. Evol. Microbiol. 0:0-0(0).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-
CC RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-
CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.

```

CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-
 CC PHOSPHO-D-GLYCERATE.
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
 CC EMBL: AJ293142; CAC51565.1; -
 DR InterPro: IPR000685; Rubisco_large.
 DR Pfam: PF00016; Rubisco_large; 1.
 DR Carbon dioxide fixation; lyase; Monooxygenase; Oxidoreductase;
 KW Photorespiration.
 FT NON_TER 1
 SQ SEQUENCE 83 AA; 9013 MW; 00F555CA851F25D CRC64;
 Query Match 1.6%; Score 8; DB 2; Length 83;
 Best Local Similarity 100.0%; Pred. No. 3.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 315 IFGDDSVL 322
 Db 1 IFGDDSVL 8
 RESULT 8
 ID Q93A03 PRELIMINARY; PRT; 83 AA.
 AC Q93A03;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO
 DE large subunit) (Fragment).
 GN RBCL.
 OS Anabaena cf. gracile 271.
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaenozomenon.
 OX NCBI_TaxID=136082;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=271;
 RA Gugger M., Lyra C., Henriksen P., Couté A., Humbert J.F., Sivonen K.;
 RT "Morphological and molecular comparison of the cyanobacterial genera
 RT Anabaena and Anabaenozomenon".
 RL Int. J. Syst. Evol. Microbiol. 0:0-0(0).
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-
 CC RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
 CC ACTIVE SITE (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-
 CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 2 3-
 CC PHOSPHO-D-GLYCERATE.
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
 CC EMBL: AJ293147; CAC51583.1; -
 DR InterPro: IPR000685; Rubisco_large.
 DR Pfam: PF00016; Rubisco_large; 1.
 DR Carbon dioxide fixation; lyase; Monooxygenase; Oxidoreductase;
 KW Photorespiration.
 FT NON_TER 1
 SQ SEQUENCE 83 AA; 9013 MW; 00F555CA851F25D CRC64;
 Query Match 1.6%; Score 8; DB 2; Length 83;
 Best Local Similarity 100.0%; Pred. No. 3.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 315 IFGDDSVL 322
 Db 1 IFGDDSVL 8
 RESULT 9
 ID Q93A00 PRELIMINARY; PRT; 83 AA.

AC Q93A00;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO
 DE large subunit) (Fragment).
 GN RBCL.
 OS Anabaena cf. cylindrica 133.
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
 OX NCBI_TaxID=136078;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=133;
 RA Gugger M., Lyra C., Henriksen P., Couté A., Humbert J.F., Sivonen K.;
 RT "Morphological and molecular comparison of the cyanobacterial genera
 RT Anabaena and Anabaenozomenon".
 RL Int. J. Syst. Evol. Microbiol. 0:0-0(0).
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-
 CC RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
 CC ACTIVE SITE (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-
 CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-
 CC PHOSPHO-D-GLYCERATE.
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
 CC EMBL: AJ293158; CAC51604.1; -
 DR InterPro: IPR000685; Rubisco_large.
 DR Pfam: PF00016; Rubisco_large; 1.
 DR Carbon dioxide fixation; lyase; Monooxygenase; Oxidoreductase;
 KW Photorespiration.
 FT NON_TER 1
 SQ SEQUENCE 83 AA; 8990 MW; 906DEC461851F24D CRC64;
 Query Match 1.6%; Score 8; DB 2; Length 83;
 Best Local Similarity 100.0%; Pred. No. 3.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 315 IFGDDSVL 322
 Db 1 IFGDDSVL 8
 RESULT 10
 ID Q939Z8 PRELIMINARY; PRT; 83 AA.
 AC Q939Z8;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO
 DE large subunit) (Fragment).
 GN RBCL.
 OS Anabaena crassa.
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
 OX NCBI_TaxID=136074;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=215;
 RA Gugger M., Lyra C., Henriksen P., Couté A., Humbert J.F., Sivonen K.;
 RT "Morphological and molecular comparison of the cyanobacterial genera
 RT Anabaena and Anabaenozomenon".
 RL Int. J. Syst. Evol. Microbiol. 0:0-0(0).
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-
 CC RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
 CC ACTIVE SITE (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-

CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-
CC PHOSPHO-D-GLYCERATE.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
DR EMBL: AJ293165; CAC51606.1; .
DR InterPro: IPR000685; RUBISCO_large.
DR Pfam: PF00016; RUBISCO_large; 1.
KM Carbon dioxide fixation; lyase; Monooxygenase; Oxidoreductase;
KW Photorespiration.
FT NON_TER 1
SQ SEQUENCE 83 AA; 9013 MW; 00F5555CA851F25D CRC64;
Query Match 1.6%; Score 8; DB 2; Length 83;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 315 IFGDDSVL 322
DB 1 IFGDDSVL 8
RESULT 11
OY 093927 PRELIMINARY; PRT; 83 AA.
AC 093927;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO large subunit) (Fragment).
GN RBCL.
OS Anabaena sp. PMC9701.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
OX NCBI_TaxID=136080;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PMC9701;
RA Guiger M., Lyra C., Henriksen P., Couté A., Humbert J.F., Sivonen K.;
RT "Morphological and molecular comparison of the cyanobacterial genera Anabaena and Aphanizomenon.";
RT Int. J. Syst. Evol. Microbiol. 0:0-0(0).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME ACTIVE SITE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-
CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-
CC PHOSPHO-D-GLYCERATE.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
DR EMBL: AJ293163; CAC51608.1; .
DR InterPro: IPR000685; RUBISCO_large.
DR Pfam: PF00016; RUBISCO_large; 1.
KM Carbon dioxide fixation; lyase; Monooxygenase; Oxidoreductase;
KW Photorespiration.
FT NON_TER 1
SQ SEQUENCE 83 AA; 9013 MW; 00F5555CA851F25D CRC64;
Query Match 1.6%; Score 8; DB 2; Length 83;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 315 IFGDDSVL 322
DB 1 IFGDDSVL 8
RESULT 12
OY 093925 PRELIMINARY; PRT; 83 AA.
AC 093925;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO large subunit) (Fragment).
GN RBCL.
OS Anabaena cf. cylindrica PMC9705.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
OX NCBI_TaxID=136079;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PMC9705;
RA Guiger M., Lyra C., Henriksen P., Couté A., Humbert J.F., Sivonen K.;
RT "Morphological and molecular comparison of the cyanobacterial genera Anabaena and Aphanizomenon.";
RT Int. J. Syst. Evol. Microbiol. 0:0-0(0).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME ACTIVE SITE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-
CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-
CC PHOSPHO-D-GLYCERATE.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
DR EMBL: AJ293165; CAC51610.1; .
DR InterPro: IPR000685; RUBISCO_large.
DR Pfam: PF00016; RUBISCO_large; 1.
KM Carbon dioxide fixation; lyase; Monooxygenase; Oxidoreductase;
KW Photorespiration.
FT NON_TER 1
SQ SEQUENCE 83 AA; 9013 MW; 00F5555CA851F25D CRC64;
Query Match 1.6%; Score 8; DB 2; Length 83;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 315 IFGDDSVL 322
DB 1 IFGDDSVL 8
RESULT 13
OY 093924 PRELIMINARY; PRT; 83 AA.
AC 093924;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO large subunit) (Fragment).
GN RBCL.
OS Aphanizomenon cf. flos-aquae PMC9501.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Aphanizomenon.
OX NCBI_TaxID=136083;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PMC9501;
RA Guiger M., Lyra C., Henriksen P., Couté A., Humbert J.F., Sivonen K.;
RT "Morphological and molecular comparison of the cyanobacterial genera Anabaena and Aphanizomenon.";
RT Int. J. Syst. Evol. Microbiol. 0:0-0(0).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME ACTIVE SITE (BY SIMILARITY).

ID 093925 PRELIMINARY; PRT; 83 AA.
AC 093925;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO large subunit) (Fragment).
GN RBCL.
OS Anabaena cf. cylindrica PMC9705.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
OX NCBI_TaxID=136079;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PMC9705;
RA Guiger M., Lyra C., Henriksen P., Couté A., Humbert J.F., Sivonen K.;
RT "Morphological and molecular comparison of the cyanobacterial genera Anabaena and Aphanizomenon.";
RT Int. J. Syst. Evol. Microbiol. 0:0-0(0).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME ACTIVE SITE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-
CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-
CC PHOSPHO-D-GLYCERATE.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
DR EMBL: AJ293165; CAC51610.1; .
DR InterPro: IPR000685; RUBISCO_large.
DR Pfam: PF00016; RUBISCO_large; 1.
KM Carbon dioxide fixation; lyase; Monooxygenase; Oxidoreductase;
KW Photorespiration.
FT NON_TER 1
SQ SEQUENCE 83 AA; 9013 MW; 00F5555CA851F25D CRC64;
Query Match 1.6%; Score 8; DB 2; Length 83;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 315 IFGDDSVL 322
DB 1 IFGDDSVL 8
RESULT 13
OY 093924 PRELIMINARY; PRT; 83 AA.
AC 093924;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO large subunit) (Fragment).
GN RBCL.
OS Aphanizomenon cf. flos-aquae PMC9501.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Aphanizomenon.
OX NCBI_TaxID=136083;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PMC9501;
RA Guiger M., Lyra C., Henriksen P., Couté A., Humbert J.F., Sivonen K.;
RT "Morphological and molecular comparison of the cyanobacterial genera Anabaena and Aphanizomenon.";
RT Int. J. Syst. Evol. Microbiol. 0:0-0(0).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME ACTIVE SITE (BY SIMILARITY).

```

CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-
CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-
CC PHOSPHO-D-GLYCERATE.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
CC EMBL: AJ293159; CAC51616.1; -.
CC InterPro: IPR000685; Rubisco_large.
DR Pfam: PF00016; Rubisco_large.
DR Carbon dioxide fixation; Lyase; Monooxygenase; Oxidoreductase;
KW Photorespiration.
FT NON_TER 1
SQ SEQUENCE 83 AA; 9013 MW; 00F5555CA851F25D CRC64;

Query Match 1.6%; Score 8; DB 2; Length 83;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 IFGDDSVL 322
DB 1 IFGDDSVL 8

RESULT 14
ID 093922 PRELIMINARY; PRT; 83 AA.
AC 093922:
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (Rubisco
DE large subunit) (Fragment).
GN RBCL
OS Anabaena macrospora.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
OX NCBI_TaxID=135103;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=PMC9301;
RA Gugger M., Lyra C., Henriksen P., Couto A., Humbert J.F., Sivonen K.;
RT "Morphological and molecular comparison of the cyanobacterial genera
RT Anabaena and Aphanizomenon.";
RL Int. J. Syst. Evol. Microbiol. 0:0-0(0).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-
CC RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-
CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-
CC PHOSPHO-D-GLYCERATE.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
CC EMBL: AJ293151; CAC51631.1; -.
CC InterPro: IPR000685; Rubisco_large.
DR Pfam: PF00016; Rubisco_large.
DR Carbon dioxide fixation; Lyase; Monooxygenase; Oxidoreductase;
KW Photorespiration.
FT NON_TER 1
SQ SEQUENCE 83 AA; 9013 MW; 00F5555CA851F25D CRC64;

Query Match 1.6%; Score 8; DB 2; Length 83;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 IFGDDSVL 322
DB 1 IFGDDSVL 8

```

RESULT 15

```

Q939W9
ID Q939W9 PRELIMINARY; PRT; 83 AA.
AC Q939W9:
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (Rubisco
DE large subunit) (Fragment).
GN RBCL.
OS Anabaena menoltae.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
OX NCBI_TaxID=136070;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=57;
RA Gugger M., Lyra C., Henriksen P., Couto A., Humbert J.F., Sivonen K.;
RT "Morphological and molecular comparison of the cyanobacterial genera
RT Anabaena and Aphanizomenon.";
RL Int. J. Syst. Evol. Microbiol. 0:0-0(0).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-
CC RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-
CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-
CC PHOSPHO-D-GLYCERATE.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
CC EMBL: AJ293155; CAC51633.1; -.
CC InterPro: IPR000685; Rubisco_large.
DR Pfam: PF00016; Rubisco_large.
DR Carbon dioxide fixation; Lyase; Monooxygenase; Oxidoreductase;
KW Photorespiration.
FT NON_TER 1
SQ SEQUENCE 83 AA; 9013 MW; 00F5555CA851F25D CRC64;

Query Match 1.6%; Score 8; DB 2; Length 83;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 IFGDDSVL 322
DB 1 IFGDDSVL 8

```

Search completed: November 29, 2002, 12:46:11
 Job time : 36 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2002, 10:10:43 ; Search time 70 seconds
(without alignments)
946.079 Million cell updates/sec

Title: US-09-868-760-7

Perfect score: 2794
Sequence: 1 DGEQNDGNCKDHDHDD.....MOGHEVESERVACVIGRA 497

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A.GeneSeq_101002:*

- 1: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT:*
- 2: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT:*
- 3: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT:*
- 4: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT:*
- 5: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT:*
- 6: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT:*
- 7: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT:*
- 8: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT:*
- 9: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT:*
- 10: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT:*
- 11: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT:*
- 12: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT:*
- 13: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT:*
- 14: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT:*
- 15: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT:*
- 16: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT:*
- 17: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT:*
- 18: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT:*
- 19: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT:*
- 20: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT:*
- 21: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT:*
- 22: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT:*
- 23: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2794	100.0	497	21	AAV93750
2	260	9.3	1529	17	AA97985
3	217	7.8	351	13	AA924393
4	173.5	6.2	244	16	AA652476
5	172.5	6.2	221	13	AA927934
6	170	6.1	221	13	AA924232
7	169.5	6.1	221	13	AA924231
8	169.5	6.1	210	13	AA924230
9	169.5	6.1	228	13	AA924227
10					GAG fusion protein

11	168	6.0	205	13	AA927935	GAG fusion protein
12	167	6.0	153	8	AA970979	Metalloprotein
13	167	6.0	154	23	AA952499	Superoxide dismutase
14	167	6.0	189	13	AA927936	GAG fusion protein
15	167	6.0	205	13	AA927932	GAG fusion protein
16	167	6.0	598	22	ABG23595	Novel human diaphorase
17	166.5	6.0	1939	22	ABG20505	Novel human diaphorase
18	165.5	5.9	153	23	AB977456	Fungal stress response
19	165.5	5.9	186	13	AA927946	GAG fusion protein
20	165.5	5.9	199	13	AA927945	GAG fusion protein
21	165.5	5.9	212	13	AA927944	GAG fusion protein
22	165.5	5.9	392	23	AA977312	HIV-1 env4/human
23	164.5	5.9	153	23	AA952495	Superoxide dismutase
24	164.5	5.9	392	23	AA952495	SOD/Env-4 fusion protein
25	164	5.9	151	23	AA952479	Superoxide dismutase
26	164	5.9	152	23	AA952489	zeta mays protein f
27	163.5	5.9	153	23	AA952489	Superoxide dismutase
28	163	5.8	152	19	AA956672	Paddy copper zinc
29	163	5.8	152	21	AA912538	GAG fusion protein
30	163	5.8	183	13	AA927948	GAG fusion protein
31	163	5.8	183	13	AA927951	GAG fusion protein
32	163	5.8	184	13	AA927941	GAG fusion protein
33	163	5.8	185	13	AA927938	GAG fusion protein
34	163	5.8	186	13	AA927943	GAG fusion protein
35	163	5.8	189	13	AA927933	GAG fusion protein
36	163	5.8	192	13	AA924229	GAG fusion protein
37	163	5.8	192	13	AA927947	GAG fusion protein
38	163	5.8	192	13	AA927950	GAG fusion protein
39	163	5.8	192	13	AA927940	GAG fusion protein
40	163	5.8	196	13	AA924225	GAG fusion protein
41	163	5.8	197	13	AA927937	GAG fusion protein
42	163	5.8	199	13	AA927942	GAG fusion protein
43	163	5.8	203	13	AA927949	GAG fusion protein
44	163	5.8	203	13	AA924235	GAG fusion protein
45	163	5.8	206	13	AA927939	GAG fusion protein

ALIGNMENTS

RESULT 1	
AAV93750	standard; Protein: 497 AA.
AAV93750:	
03-OCT-2000 (first entry)	
AAV93750:	
Amino acid sequence of a serine protease inhibitor protein.	
Serine protease inhibitor; green-lipped mussel; anti-Chrombin;	
divalent metal cation binding activity; dietary supplement;	
anticoagulant.	
OS	Perna canaliculus.
PN	WO200039165-A1.
PD	06-JUL-2000.
PF	23-DEC-1999; 99WO-N200227.
PR	23-DEC-1998; 98NZ-0333568.
PT	23-JUL-1999; 99NZ-0336906.
PA	(HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.
PI	Scotti PD, Dearing SC, Greenwood DR, Newcomb RD;
DR	WPI: 2000-452375/39.
DR	N-PSDB: AAA47150, AAA47151.
PT	New Perna canaliculus serine protease inhibitor protein exhibiting


```

RESULT 3
ID AAR24393 standard; Protein; 351 AA.
XX
AC AAR24393;
XX
DT 22-NOV-1992 (first entry)
XX
DE Sequence of Histidine-rich protein (HISRP).
XX
KW Malaria vaccine; Histidine-rich protein; cytoadherence.
XX
OS Plasmodium lophurae.
XX
FT Key Location/Qualifiers
FT Modified-site 40..42
FT Peptide /label= potential glycosylation site
FT Peptide 1..23
FT Peptide /label= signal
FT Peptide 24..47
FT Peptide /label= pro-peptide
XX
UN US5116965-A.
XX
PD 25-MAY-1992.
XX
PE 26-AUG-1986; 86US-0900401.
XX
PR 26-AUG-1986; 86US-0900401.
PA (SILOK ) SILOAN KETTERING INST CANCER.
PI Pologe L.; Ravetch JV;
XX
DR WPI: 1992-199590/24.
DR N-PSDB; AAQ24393.
XX
PT Histidine-rich protein associated with Plasmodium knob phenotype -
PT and DNA encoding it, used for in vitro diagnosis of P.
PT falciparum infection.
XX
PS Disclosure; Fig 7A-B; 29pp; English.
XX
CC Two variants of HISRP are produced by P. falciparum. One is
CC associated with what is referred to as "knobby phenotype" (K30) and
CC "knobless phenotype" (K-). The "Knobby" and "knobless" phenotypes
CC have been implicated in cytoadherence, which is characteristic of
CC erythrocyte infection. It has now been found that cDNA expressing
CC both K+ and K- HISRP can be obtained by the use of P. lophurae HISRP
CC expressing DNA. The genomic clone (AAQ25532) is encoded in two
CC exons, separating the signal peptide-encoding sequence from the
CC pro-sequence, confirming that synthesis of the protein occurs via
CC the preproprotein. Oligo. probes synthesised to the signal
CC peptide-encoding exon reveal multiple homologous DNA sequences in
CC the P. lophurae genome. The sequence of mature proteins is arranged
CC in numerous tandem repeats with up to nine histidine residues in a
CC row, similar to other Plasmodium proteins for which sequence data
CC have so far been reported.
XX
SO Sequence 351 AA;
XX
Query Match 7.8%; Score 217; DB 13; Length 351;
Best Local Similarity 19.7%; Pred. No. 8.6e-12;
Matches 84; Conservative 29; Mismatches 164; Indels 150; Gaps 17.
OY 12 DDHHDDNDHDNDHDDDDDELNTHTAO-CSEMRPNMSSLLNNHYNGSTLSOKGCGAVYL 70
OY :::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
OY 67 EENERRNEENRNEENRNEENRNEENRNEENRNEENRNEENRNEENRNEENRNEENRNEENR 120
OY ELRLRVGNESDHDNDHGLNLMLSDMSAGCSDSIGELYNNAPEKNADPDLADLYDDOR 130

```

[illegible]

PS Claim 1: Page 24-25: 55pp; English.
XX
XX A rat C6 glioma (ATCC CCL 107) cDNA library in phage lambda gtl1 was
CC screened with a cDNA probe based on the partial amino acid sequence
CC of rat extracellular superoxide-dismutase form B (EC-SOD-B) to
CC isolate the full-size rat EC-SOD cDNA given in AA075336. Recombinant
CC EC-SOD-B polypeptides have been produced for therapeutic
CC applications.

XX Sequence 244 AA:

Query Match 6.2%; Score 173.5; DB 16; Length 244;
Best Local Similarity 29.7%; Pred. No. 6.3e-08;
Matches 54; Conservative 28; Mismatches 61; Indels 39; Gaps 10;

194 LHHLEEDKTEHYACDVRSNTHOKAL---HHHVGITDFKQVGYD-LEVSYHLEGF 248
55 LGKQREADAREMHAACRY-----QPSAMLPDPOQITGLVLFRLQSPSSRLSEAFNLEGF 109
249 NVSDDKHDLHDVOIYANGDLTSGCDNLGAKYDPHEDYHSELGDLGDIDDDHG--VYNE 306
110 PABQNTSNHA--IHVHEFGDLSGCESTGPHYNP-----LGVPHPQHPDGFNFVARD 160
307 ----SHRYSW-INIFGDDSVLGRSIAIHQ-RDHLAKSA-----KIACCVYGRG 348
161 GRMKHMGSLATSLAGPHSILGRAVYHAGEDDLGKGNQASVQNGAGRRRLACCVYGRS 220
349 QS 350
221 NS 222

RESULT 5

AAM52476
ID AAM52476 standard; Protein; 244 AA.

XX AAM52476;

XX 03-JUL-2002 (first entry)

XX Superoxide dismutase protein #5.

XX Selenoprotein; HIV; Ebola virus; cancer; immune system disorder.

XX Rattus sp.

XX US6303295-B1.

XX 16-OCT-2001.

XX 12-JUL-1996; 96US-0679493.

XX 14-JUL-1995; 95US-001203P.

XX 01-SEP-1995; 95US-003112P.

XX (UYGE-) UNIV GEORGIA RES FOUND INC.

XX Taylor EW, Nadimpalli RG, Ramanathan CS;

XX WPI: 2002-024734/03.

XX New selenoprotein for use in detecting certain viruses, e.g. human
XX immunodeficiency virus (HIV) or Ebola, cancer and immune system
XX disorders -

XX Disclousure; Columns 77-88: 140pp; English.

XX The present invention relates to selenoproteins encoded in the genome of
XX a virus, where the coding sequence of the selenoprotein is genetically
XX engineered for expression in a nucleic acid construct. The invention also
XX discloses a method for identifying selenoprotein coding sequences, for
XX detecting certain viruses (e.g. HIV or Ebola), cancer and immune system
XX disorders. The present sequence was used to illustrate the invention.

XX Sequence 244 AA:

Query Match 6.2%; Score 173.5; DB 23; Length 244;
Best Local Similarity 29.7%; Pred. No. 6.3e-08;
Matches 54; Conservative 28; Mismatches 61; Indels 39; Gaps 10;

194 LHHLEEDKTEHYACDVRSNTHOKAL---HHHVGITDFKQVGYD-LEVSYHLEGF 248
55 LGKQREADAREMHAACRY-----QPSAMLPDPOQITGLVLFRLQSPSSRLSEAFNLEGF 109
249 NVSDDKHDLHDVOIYANGDLTSGCDNLGAKYDPHEDYHSELGDLGDIDDDHG--VYNE 306
110 PABQNTSNHA--IHVHEFGDLSGCESTGPHYNP-----LGVPHPQHPDGFNFVARD 160
307 ----SHRYSW-INIFGDDSVLGRSIAIHQ-RDHLAKSA-----KIACCVYGRG 348
161 GRMKHMGSLATSLAGPHSILGRAVYHAGEDDLGKGNQASVQNGAGRRRLACCVYGRS 220
349 QS 350
221 NS 222

RESULT 6

AAR27934
ID AAR27934 standard; Protein; 221 AA.

XX AAR27934;

XX 25-NOV-1992 (first entry)

XX GAG fusion protein with SOD according to a formula.

XX glycosamino:glycan; superoxidizedismutase; tissue damage;

XX autoimmune disease; rheumatoid arthritis; osteoarthritis; ss.

XX Synthetic.

XX Key Location/Qualifiers

XX Peptide 1..172

XX Peptide 173..221

XX Peptide 173..221

XX WO9207935-A.

XX 14-MAY-1992.

XX 01-NOV-1991; 91WO-US08105.

XX 01-NOV-1990; 90US-0608539.

XX 02-NOV-1990; 90US-0608569.

XX (SCRI) SCRIPPS RES INST.

XX Boltsfnoct M, Fisher C, Griffen JH, Hallowell RA, Kuhn L;

XX Mullendachgt, Parge HE, Tainer JA;

XX WPI: 1992-183671/22.

XX Claim 8; Fig 1; 140pp; English.

XX The fusion protein comprising the a glycosaminoglycan binding region
XX and human superoxide dismutase, joined via a linker region was
XX constructed according to the formula SOD-(M-2/3)-M where 2 is the peptide
XX -HNNHREKKKRVEDL-. The fusion protein is useful for extending
XX the in vivo lifetimes of biologically active cpds. such as SOD and

for targeting them to specific cell surfaces or substrates. The glycosaminoglycan (GAG) binding protein is formed into a fusion protein with SOD to increase stability, plasma half-life and ease of purification of SOD. SOD is useful for reduction of tissue damage caused by oxygen radicals and is used in the treatment of autoimmune diseases e.g. rheumatoid and osteo arthritis.

See also AAR24225-35, AAR27932-51.

Sequence 221 AA:

Query Match 6.28; Score 172.5; DB 13; Length 221;

Best Local Similarity 23.58; Pred. No. 6.8e-08;

Matches 60; Conservative 40; Mismatches 72; Indels 83; Gaps 10;

225 VHGTFDKO-VGYGDELYSYHLEGFNVSDDKHDLHVOIYANGDLTSGCDNLGAKYDP- 282

15 VQGIINFEDKESNGPVKWKWSIKGLT-----EGLHGFHNEHGDVMTAGCTSGPFPNPL 68

283 -----HEDYSELGLDLDIHDDHGVNESHRYSWINIFGDDSVLGRSLAHOR-DHLH 335

69 SRKHGGRPKDERHNVGLDGNVTADKGVADVSIEDSVISLGSDDHSIIIGRTLVYHEKADDLG 128

336 K-----SAKIACVYIGRQGS-----HP-----EIVHRACVVRPNTESTG 370

129 KGNBESTKTGNAGSRILACGVIGIAGCGHHRHREMKKRVEDLMHR----- 175

371 LHHNVSGSITFEQTPGSGTHTADLKGFNVSELSHHRHGVOLHEMGDSHGCHSLGRMY 430

176 -----HHRPREK-KRVEDLMHRHHRPREMK-----RVEDLM 205

431 HGHDAHDKRPDGL 445

206 HRRHPREMKRVEDL 220

RESULT 7

AAR24232

ID AAR24232 standard; Protein: 221 AA.

AC AAR24232;

25-NOV-1992 (first entry)

GAG fusion protein with SOD according to a formula.

Glycosamino:glycan; superoxidisedismutase; tissue damage;

autoimmune disease; rheumatoid arthritis; osteoarthritis; ss.

Synthetic.

Key Location/Qualifiers

Peptide 1..49

/note= "GAG binding motif"

50..221

/note= "SOD"

WO9207935-A.

14-MAY-1992.

01-NOV-1991; 91WO-US08105.

01-NOV-1990; 90US-0608539.

02-NOV-1990; 90US-0608569.

(SCRT) SCRIPPS RES INST.

Boisjout M, Fisher C, Griffin JH, Hallewell RA, Kuhn L;

Mullenbachgt, Pargie HE, Tainer JA;

WPI: 1992-183671/22.

Fusion proteins with glycosamino:glycan-binding and

PT superoxidisedismutase activities - reduce tissue damage caused by

PT superoxide radicals, useful in treating autoimmune diseases e.g.

PT rheumatoid arthritis and osteoarthritis

PS Claim 8; Fig 1: 140pp; English.

XX The fusion protein comprising the a glycosaminoglycan binding region

CC and human superoxide dismutase, joined via a linker region was

CC constructed according to the formula M-(Z-M)3-SOD where Z is the peptide

CC -HRRHPREMKRVEDL-. The fusion protein is useful for extending

CC the in vivo lifetimes of biologically active cps. such as SOD and

CC glycosaminoglycan (GAG) binding protein is formed into a fusion

CC protein with SOD to increase stability, plasma half-life and ease

CC of purification of SOD. SOD is useful for reduction of tissue damage

CC caused by oxygen radicals and is used in the treatment of autoimmune

CC diseases e.g. rheumatoid and osteo arthritis.

See also AAR24225-35, AAR27932-51.

PT superoxidisedismutase activities - reduce tissue damage caused by

PT superoxide radicals, useful in treating autoimmune diseases e.g.

PT rheumatoid arthritis and osteoarthritis

PS Claim 8; Fig 1: 140pp; English.

XX The fusion protein comprising the a glycosaminoglycan binding region

CC and human superoxide dismutase, joined via a linker region was

CC constructed according to the formula M-(Z-M)3-SOD where Z is the peptide

CC -HRRHPREMKRVEDL-. The fusion protein is useful for extending

CC the in vivo lifetimes of biologically active cps. such as SOD and

CC glycosaminoglycan (GAG) binding protein is formed into a fusion

CC protein with SOD to increase stability, plasma half-life and ease

CC of purification of SOD. SOD is useful for reduction of tissue damage

CC caused by oxygen radicals and is used in the treatment of autoimmune

CC diseases e.g. rheumatoid and osteo arthritis.

See also AAR24225-35, AAR27932-51.

Sequence 221 AA:

Query Match 6.18; Score 170; DB 13; Length 221;

Best Local Similarity 25.78; Pred. No. 1.2e-07;

Matches 55; Conservative 34; Mismatches 67; Indels 58; Gaps 9;

185 KAPETAAALHHELE-EDKTEHYAHCDVRSNTROPKALHH----- 224

10 KRVEDLMHRHHRPREMKRVEDLMH-----RHHRPREMKRVEDLMATKAVAVLKGDGP 63

225 VHGTFDKO-VGYGDELYSYHLEGFNVSDDKHDLHVOIYANGDLTSGCDNLGAKYDP- 282

64 VQGIINFEDKESNGPVKWKWSIKGLT-----EGLHGFHNEHGDVMTAGCTSGPFPNPL 117

283 -----HEDYSELGLDLDIHDDHGVNESHRYSWINIFGDDSVLGRSLAHOR-DHLH 335

118 SRKHGGRPKDERHNVGLDGNVTADKGVADVSIEDSVISLGSDDHSIIIGRTLVYHEKADDLG 177

336 K-----SAKIACVYIGRQGS-----HP 352

178 KGNBESTKTGNAGSRILACGVIGIAGCGHHRHHP 211

RESULT 8

AAR24231

ID AAR24231 standard; Protein: 192 AA.

AC AAR24231;

25-NOV-1992 (first entry)

GAG fusion protein with SOD according to a formula.

Glycosamino:glycan; superoxidisedismutase; tissue damage;

autoimmune disease; rheumatoid arthritis; osteoarthritis; ss.

Synthetic.

Key Location/Qualifiers

Peptide 1..172

/note= "SOD"

173..192

/note= "GAG binding motif"

WO9207935-A.

14-MAY-1992.

01-NOV-1991; 91WO-US08105.

01-NOV-1990; 90US-0608539.

02-NOV-1990; 90US-0608569.

(SCRT) SCRIPPS RES INST.

```
XX Bolstein M, Fisher C, Griffin JH, Hallelwell RA, Kuhn L;  
PI Mullenbachgt, Parge HE, Tainer JA;  
XX WPI: 1992-183671/22.  
XX  
XX Fusion proteins with glycosamino:glycan-binding and  
PT superoxidismutase activities - reduce tissue damage caused by  
PT superoxide radicals; useful in treating autoimmune diseases e.g.  
PT rheumatoid arthritis and osteoarthritis  
XX  
XX Claim 7; Fig 1; 140pp: English.  
XX  
XX The fusion protein comprising the a glycosaminoglycan binding region  
CC and human superoxide dismutase, joined via a linker region was  
CC constructed according to the formula SOD-(M-2)-M where 2 is the peptide  
CC -RPRSGKKRRKKRLKRS-. The fusion protein is useful for extending  
CC the in vivo lifetimes of biologically active epds, such as SOD and  
CC for targeting them to specific cell surfaces or substrates. The  
CC glycosaminoglycan (GAG) binding protein is formed into a fusion  
CC protein with SOD to increase stability, plasma half-life and ease  
CC of purification of SOD. SOD is useful for reduction of tissue damage  
CC caused by oxygen radicals and is used in the treatment of autoimmune  
CC diseases e.g. rheumatoid and osteo-arthritis.  
CC See also AAR24225-35, AAR27932-51.  
XX  
XX Sequence 192 AA:  
SQ  
Query Match 6.1%; Score 169.5; DB 13; Length 192;  
Best Local Similarity 29.2%; Pred. No. 1.1e-07;  
Matches 50; Conservative 34; Mismatches 54; Indels 33; Gaps 7;  
OY 225 VHGTFIDFKQ-VGYGDLEVSYLEGFNVSDDKHDLHDVQIYANGDLTSGCDNLGAKYDP-282  
DB 15 VQGLINEQKESNGPVKVGWSIKGLT-----EGLHGPHVHEFGDNTAGCTSAQPHFNPL 68  
OY 283 -----HEDYHSELGDLGDIHDDHGVNESHRYSMINIFGDDSVLGRSAIAHOR-DHLH 335  
DB 69 SRKHGKPRDEERHNGDIGNVTADKGDVADVSIEDSVISLSCDHSITIGRTLVVHEKADLDG 128  
OY 336 K-----SAKIACCVIGRGOS-----HP-EIYHRAKCVVRPTES 368  
DB 129 KGNNESTKTGNAGSRLACGVIGIAGPGRHNRHPRMKKRVEDLMRVPRRES 179  
RESULT 9  
AAR24230 standard; Protein: 210 AA.  
XX AAR24230;  
XX  
XX 25-NOV-1992 (first entry)  
XX  
XX GAG fusion protein with SOD according to a formula.  
XX  
XX Glycosamino:glycan: superoxidismutase; tissue damage;  
XX autoimmune disease; rheumatoid arthritis; osteoarthritis; ss.  
XX  
XX Synthetic.  
XX  
XX Key Location/Qualifiers  
FH Peptide 1..172  
FT /note- "SOD"  
FT 173..210  
FT Peptide /note- "GAG binding motif"  
XX  
XX W09207935-A.  
XX  
XX 14-MAY-1992.  
XX  
XX 01-NOV-1991; 91MO-US08105.  
XX  
XX 01-NOV-1990; 90US-0608539.  
XX
```

```
PR 02-NOV-1990; 90US-0608569.  
XX  
XX (SCRI.) SCRIIPS RES INST.  
XX  
XX Bolstein M, Fisher C, Griffin JH, Hallelwell RA, Kuhn L;  
PI Mullenbachgt, Parge HE, Tainer JA;  
XX WPI: 1992-183671/22.  
XX  
XX Fusion proteins with glycosamino:glycan-binding and  
PT superoxidismutase activities - reduce tissue damage caused by  
PT superoxide radicals; useful in treating autoimmune diseases e.g.  
PT rheumatoid arthritis and osteoarthritis  
XX  
XX Claim 7; Fig 1; 140pp: English.  
XX  
XX The fusion protein comprising the a glycosaminoglycan binding region  
CC and human superoxide dismutase, joined via a linker region was  
CC constructed according to the formula SOD-(M-2)-M where 2 is the peptide  
CC -RPRSGKKRRKKRLKRS-. The fusion protein is useful for extending  
CC the in vivo lifetimes of biologically active epds, such as SOD and  
CC for targeting them to specific cell surfaces or substrates. The  
CC glycosaminoglycan (GAG) binding protein is formed into a fusion  
CC protein with SOD to increase stability, plasma half-life and ease  
CC of purification of SOD. SOD is useful for reduction of tissue damage  
CC caused by oxygen radicals and is used in the treatment of autoimmune  
CC diseases e.g. rheumatoid and osteo-arthritis.  
CC See also AAR24225-35, AAR27932-51.  
XX  
XX Sequence 210 AA:  
SQ  
Query Match 6.1%; Score 169.5; DB 13; Length 210;  
Best Local Similarity 29.2%; Pred. No. 1.2e-07;  
Matches 50; Conservative 34; Mismatches 54; Indels 33; Gaps 7;  
OY 225 VHGTFIDFKQ-VGYGDLEVSYLEGFNVSDDKHDLHDVQIYANGDLTSGCDNLGAKYDP-282  
DB 15 VQGLINEQKESNGPVKVGWSIKGLT-----EGLHGPHVHEFGDNTAGCTSAQPHFNPL 68  
OY 283 -----HEDYHSELGDLGDIHDDHGVNESHRYSMINIFGDDSVLGRSAIAHOR-DHLH 335  
DB 69 SRKHGKPRDEERHNGDIGNVTADKGDVADVSIEDSVISLSCDHSITIGRTLVVHEKADLDG 128  
OY 336 K-----SAKIACCVIGRGOS-----HP-EIYHRAKCVVRPTES 368  
DB 129 KGNNESTKTGNAGSRLACGVIGIAGPGRHNRHPRMKKRVEDLMRVPRRES 179  
RESULT 10  
AAR24227 standard; Protein: 228 AA.  
XX AAR24227;  
XX  
XX 25-NOV-1992 (first entry)  
XX  
XX GAG fusion protein with SOD according to a formula.  
XX  
XX Glycosamino:glycan: superoxidismutase; tissue damage;  
XX autoimmune disease; rheumatoid arthritis; osteoarthritis; ss.  
XX  
XX Synthetic.  
XX  
XX Key Location/Qualifiers  
FH Peptide 1..172  
FT /note- "SOD"  
FT 173..228  
FT Peptide /note- "GAG binding motif"  
XX  
XX W09207935-A.  
XX  
XX 14-MAY-1992.  
XX
```

PF 01-NOV-1991: 91WO-US08105.
 XX
 XX 01-NOV-1990: 90US-0608539.
 PR 02-NOV-1990: 90US-0608569.
 XX
 PA (SCRI) SCRIPPS RES INST.
 PI Boissinot M, Fisher C, Griffin JH, Hallewell RA, Kuhn L;
 PI Mullenbachgt, Parge HE, Tainer JA;
 DR WPI: 1992-183671/22.
 XX
 XX Fusion proteins with glycosamino:glycan-binding and
 PT superoxidismutase activities - reduce tissue damage caused by
 PT superoxide radicals, useful in treating autoimmune diseases e.g.
 PT rheumatoid arthritis and osteoarthritis
 PS Claim 7: Fig 1: 140pp; English.
 XX
 XX The fusion protein comprising the a glycosaminoglycan binding region
 CC and human superoxide dismutase, joined via a linker region was
 CC constructed according to the formula SOD-(M-Z)₃-M where Z is the peptide
 CC RNPESGKRRKRRKPS-. The fusion protein is useful for extending
 CC the in vivo lifetimes of biologically active cpds. such as SOD and
 CC for targeting them to specific cell surfaces or substrates. The
 CC glycosaminoglycan (GAG) binding protein is formed into a fusion
 CC protein with SOD to increase stability, plasma half-life and ease
 CC of purification of SOD. SOD is useful for reduction of tissue damage
 CC caused by oxygen radicals and is used in the treatment of autoimmune
 CC diseases e.g. Rheumatoid and osteo-arthritis.
 CC See also AAR24225-35, AAR27932-51.
 CC
 XX Sequence 228 AA:
 SQ
 Query Match 6.1%; Score 169.5; DB 13; Length 228;
 Best Local Similarity 29.2%; Pred. No. 1.4e-07;
 Matches 50; Conservative 34; Mismatches 54; Indels 33; Gaps 7;
 OY 225 VHGTFPKO-VGYGDELYSHLEGFVNSDDKHLDVQIYANDLISGCDNIGAKYDP- 282
 DB 15 VQGIINFEQKESNGPYKVGSIKGLT-----EGHGHVHEFGDNTAGCTSGAPHNPL 68
 OY 283 -----HEDYHSELGLDGIHDDHGVNESHRSYMIIFGDDSVLGRSAIAHQ-R-DHLH 335
 DB 69 SRKHGPKDERHVDLGNVTADKGDVADVSIEDSVISLSGDHSIIIGRTLVHEKADDLG 128
 OY 336 K-----SAKIACCVIGRQS-----HP-ETVHRAKCVRPNTESG 368
 DB 129 KGNESSTKTGNAGSRLACGVIGIAOGPGCHRRHPREMKRRVEDLMRPRES 179
 DB
 RESULT 11
 AAR27935
 ID AAR27935 standard; Protein: 205 AA.
 AC AAR27935;
 XX
 XX 25-NOV-1992 (first entry)
 DE GAG fusion protein with SOD according to a formula.
 XX
 XX glycosamino:glycan; superoxidismutase; tissue damage;
 KW autoimmune disease; Rheumatoid arthritis; osteoarthritis; ss.
 XX
 XX Synthetic.
 XX
 XX Key Location/Qualifiers
 FT Peptide 1..172
 FT /note="SOD"
 FT Peptide 173..205
 FT /note="GAG binding motif"
 FT
 PN MO9207935-A.

XX
 PD 14-MAY-1992.
 XX
 XX 01-NOV-1991: 91WO-US08105.
 PF 02-NOV-1990: 90US-0608539.
 PR 02-NOV-1990: 90US-0608569.
 XX
 PA (SCRI) SCRIPPS RES INST.
 PI Boissinot M, Fisher C, Griffin JH, Hallewell RA, Kuhn L;
 PI Mullenbachgt, Parge HE, Tainer JA;
 DR WPI: 1992-183671/22.
 XX
 XX Fusion proteins with glycosamino:glycan-binding and
 PT superoxidismutase activities - reduce tissue damage caused by
 PT superoxide radicals, useful in treating autoimmune diseases e.g.
 PT rheumatoid arthritis and osteoarthritis
 PS Claim 8: Fig 1: 140pp; English.
 XX
 XX The fusion protein comprising the a glycosaminoglycan binding region
 CC and human superoxide dismutase, joined via a linker region was
 CC constructed according to the formula SOD-(M-Z)₃-M where Z is the peptide
 CC -RRHPRREKRRKVEDL-. The fusion protein is useful for extending
 CC the in vivo lifetimes of biologically active cpds. such as SOD and
 CC for targeting them to specific cell surfaces or substrates. The
 CC glycosaminoglycan (GAG) binding protein is formed into a fusion
 CC protein with SOD to increase stability, plasma half-life and ease
 CC of purification of SOD. SOD is useful for reduction of tissue damage
 CC caused by oxygen radicals and is used in the treatment of autoimmune
 CC diseases e.g. Rheumatoid and osteo-arthritis.
 CC See also AAR24225-35, AAR27932-51.
 CC
 XX Sequence 205 AA:
 SQ
 Query Match 6.0%; Score 168; DB 13; Length 205;
 Best Local Similarity 24.7%; Pred. No. 1.6e-07;
 Matches 54; Conservative 35; Mismatches 56; Indels 74; Gaps 9;
 OY 225 VHGTFPKO-VGYGDELYSHLEGFVNSDDKHLDVQIYANDLISGCDNIGAKYDP- 282
 DB 15 VQGIINFEQKESNGPYKVGSIKGLT-----EGHGHVHEFGDNTAGCTSGAPHNPL 68
 OY 283 -----HEDYHSELGLDGIHDDHGVNESHRSYMIIFGDDSVLGRSAIAHQ-R-DHLH 335
 DB 69 SRKHGPKDERHVDLGNVTADKGDVADVSIEDSVISLSGDHSIIIGRTLVHEKADDLG 128
 OY 336 K-----SAKIACCVIGRQS-----HP-----ETVHRAKCVRPNTESG 370
 DB 129 KGNESSTKTGNAGSRLACGVIGIAOGPGCHRRHPREMKRRVEDLMHR----- 175
 OY 371 LHHVSGSITFEQTPGSGTHTADLKGFVNSDDLSHNRH 409
 DB 176 -----HHPREMK--KRVEDLMHRRH 193
 DB
 RESULT 12
 AAP70979
 ID AAP70979 standard; Protein: 153 AA.
 AC AAP70979;
 XX
 XX 03-MAY-1991 (first entry)
 DE Metalfree apoprotein for metalloprotein-albumin conjugate (II).
 XX
 XX Metalloprotein; apoprotein; albumin; antibody; antilysoin.
 KW
 XX Key Location/Qualifiers
 FT Disulfide-bond 57..146
 FT
 XX

```
PN DE3718476-A.
XX
PD 03-DEC-1987.
XX
PF 02-JUN-1987; 87DE-3718476.
XX
PR 02-JUN-1986; 86GB-0013315.
XX
PA (DEBI-) DEBIOPHARM SA.
XX
PI Deghenghl R;
XX
PS WPI: 1987-343120/49.
XX
PT Biologically active metalloprotein albumin conjugates - produced
XX by crosslinking a metal-free apoprotein in the presence of
XX albumin and adding metal to the purified conjugate
XX
XX Claim 3; Page 1; 5pp; German.
XX
XX To Al may be attached H or acetyl and the protein is obtainable by
XX the procedure of US4340675.
XX CC The protein is crosslinked in the presence of human albumin and opt.
XX an antibody fragment. The metal-free conjugate is purified, and the
XX biological activity is restored by adding the desired metal, eg
XX copper-zinc, to the conjugate. The opt. antibody is anti-lysozyme
XX material. The product enables the metalloprotein to be directed
XX to a particular tissue or organ for therapeutic purposes, the
XX immunogenicity and antigenicity to be reduced, and/or the
XX bioavailability and therapeutic activity to be increased.
XX CC The M1-albumin-antilysozyme conjugates are esp. suitable for directing
XX MP to myosin, a protein of heart muscle.
XX See also AAP70978.
XX
XX Sequence 153 AA;
XX
XX Query Match 6.0%; Score 167; DB 8; Length 153;
XX Best Local Similarity 30.1%; Pred. No. 1.3e-07;
XX Matches 43; Conservative 26; Mismatches 48; Indels 26; Gaps 5;
XX
XX QY 225 VHGTTDFKQVGYGD-LEVSYLEGFFVNSDDHKDHLHDVQIYANGDLTSGCDNLGAKYDPH 283
XX | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
XX DB 13 VSGVYKFEQASESEPTVSYEINGNSPNAERGFHIEF-----GDATNGCVSAGPHENPF 67
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX QY 284 EDYHS-----ELGDLGDIHDDHGVNESHRYSWINIFGDDSVLGRSIAIHO-RDHLH 335
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX DB 68 KKTGAPPTDEVRRHVGDGKVNKTDENGVAKGSFDSLKLIGPTSVVGRSVYIHAGDDDLG 127
XX
XX QY 336 K-----SAKIACCVIG 346
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX DB 128 KGDTEESLKTGNAGRPAPCGVIG 150
XX
XX RESULT 13
XX AAM52499
XX ID AAM52499 standard; Protein; 154 AA.
XX
XX AC AAM52499;
XX
XX DT 03-JUL-2002 (first entry)
XX
XX DE Superoxide dismutase protein #28.
XX
XX KM Selenoprotein; HIV; Ebola virus; cancer; immune system disorder.
XX
XX OS Saccharomyces cerevisiae.
XX
XX PN US6303295-B1.
XX
XX PD 16-OCT-2001.
XX
XX PF 12-JUL-1996; 96US-0679493.
XX
```

```
PR 14-JUL-1995; 95US-001203P.
PR 01-SEP-1995; 95US-003112P.
XX
XX (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
XX PI Taylor EM, Nadimpalli RG, Ramanathan CS;
XX
XX DR WPI: 2002-024734/03.
XX
XX PT New selenoprotein for use in detecting certain viruses, e.g. human
XX immunodeficiency virus (HIV) or Ebola, cancer and immune system
XX disorders -
XX
XX PS Disclosure; Columns 77-88; 140pp; English.
XX
XX CC The present invention relates to selenoproteins encoded in the genome of
XX a virus, where the coding sequence of the selenoprotein is genetically
XX engineered for expression in a nucleic acid construct. The invention also
XX discloses a method for identifying selenoprotein coding sequences, for
XX detecting certain viruses (e.g. HIV or Ebola), cancer and immune system
XX disorders. The present sequence was used to illustrate the invention.
XX
XX Sequence 154 AA;
XX
XX Query Match 6.0%; Score 167; DB 23; Length 154;
XX Best Local Similarity 30.1%; Pred. No. 1.3e-07;
XX Matches 43; Conservative 26; Mismatches 48; Indels 26; Gaps 5;
XX
XX QY 225 VHGTTDFKQVGYGD-LEVSYLEGFFVNSDDHKDHLHDVQIYANGDLTSGCDNLGAKYDPH 283
XX | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
XX DB 14 VSGVYKFEQASESEPTVSYEINGNSPNAERGFHIEF-----GDATNGCVSAGPHENPF 68
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX QY 284 EDYHS-----ELGDLGDIHDDHGVNESHRYSWINIFGDDSVLGRSIAIHO-RDHLH 335
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX DB 69 KKTGAPPTDEVRRHVGDGKVNKTDENGVAKGSFDSLKLIGPTSVVGRSVYIHAGDDDLG 128
XX
XX QY 336 K-----SAKIACCVIG 346
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX DB 129 KGDTEESLKTGNAGRPAPCGVIG 151
XX
XX RESULT 14
XX AAR27936
XX ID AAR27936 standard; Protein; 189 AA.
XX
XX AC AAR27936;
XX
XX DT 25-NOV-1992 (first entry)
XX
XX DE GAG fusion protein with SOD according to a formula.
XX
XX KW Glycosaminoglycan; superoxidizedismutase; tissue damage;
XX autoimmune disease; rheumatoid arthritis; osteoarthritis; ss.
XX
XX OS Synthetic.
XX
XX FH Key Location/Qualifiers
XX FT Peptide 1..172
XX FT /note="SOD"
XX FT Peptide 173..189
XX FT /note="GAG binding motif"
XX
XX PN W09207935-A.
XX
XX PD 14-MAY-1992.
XX
XX PF 01-NOV-1991; 91WO-US08105.
XX
XX PR 01-NOV-1990; 90US-0608539.
XX
XX PR 02-NOV-1990; 90US-0608569.
XX
XX PA (SCRI) SCRIPPS RES INST.
XX
```


P1	Boltsnagt M., Fisher C., Griffin JH, Halliwell RA, Kuhn L;
P2	Mullenbachgt, Parge HE, Tainer JA;
DR	WPI; 1992-183671/22.
XX	
PT	Fusion proteins with glycosamino:glycan-binding and
PT	superoxidisedismutase activities - reduce tissue damage caused by
PT	peroxide radicals' useful in treating autoimmune diseases e.g.
PT	rheumatoid arthritis and osteoarthritis
XX	
PS	Claim 8; Fig 1; 140pp; English.
XX	
CC	The fusion protein comprising the a glycosaminoglycan binding region
CC	and human superoxide dismutase, joined via a linker region was
CC	constructed according to the formula SOD-(M-Z)-X where Z is the peptide
CC	-NHHRPREKKRVEDEL-. The fusion protein is useful for extending
CC	the in vivo lifetimes of biologically active cpds. such as SOD and
CC	for targeting them to specific cell surfaces or substrates. The
CC	glycosaminoglycan (GAG) binding protein is formed into a fusion
CC	protein with SOD to increase stability, plasma half-life and ease
CC	of purification of SOD. SOD is useful for reduction of tissue damage
CC	caused by oxygen radicals and is used in the treatment of autoimmune
CC	diseases e.g. Rheumatoid and Osteo-arthritis.
CC	See also AAR24225-35, AAR27932-51.
XX	
SQ	Sequence 189 AA;
OQuery Match	6.0%; Score 167; DB 13; Length 189;
Best Local Similarity	28.1%; Pred. No. 1,Re-07;
Matches 47; Conservative 33; Mismatches 47; Indels 40; Gaps	
OY	225 VHCITDFKQ-VGVCGDLESYVHEGFNVSDHKHDHVDQIANDLISGCDNLGAKYDP- 282
Db	15 VQGLINEEQEKESNPVKVMGSIKGLT-----EGLHGCHVFHFEDNTAGCTSGAPHPNPL 68
OY	283 -----HEDHSELGDGLDHHDDHGYVNESKRYSWIIFGDSVLGRSTAIRQR-DLHL 335
Db	69 SRKIGGFKDERHVGDDLGAVTAADKQADVATIEDSVISLSGDHSIIIGRTIVHEKADDGA 128
OY	336 K-----SAKIACCVIGRGOS---HP-----ELVHR 357
Db	129 KGNEESTKTGNAGSRLLACGVIGIAOGPEHRHHPREMKRKREVEDLMHR 175
RESULT 15	
AAR27932	
ID	AAR27932 standard; Protein; 205 AA.
XX	
AC	AAR27932:
XX	
DT	25-NOV-1992 (first entry)
DE	GAG fusion protein with SOD according to a formula.
KW	glycosamino:glycan; superoxidisedismutase; tissue damage;
RW	autoimmune disease; rheumatoid arthritis; osteoarthritis; ss.
XX	
OS	Synthetic.
XX	
FT	Key Location/Qualifiers
FT	Peptide 1..32
FT	/note= "GAG binding motif"
FT	Peptide 33..205
FT	/note= "SOD "
XX	
PM	MO9207935-A.
PD	14-MAY-1992.
XX	
PF	01-NOV-1991; 91WO-US08105.
XX	
PR	01-NOV-1990; 90US-0608539.
PR	02-NOV-1990; 90US-0608569.

```

XX PA (SCRI ) SCRIIPS RES INST.
XX PI Boissinot M, Fisher C, Griffin JH, Hallowell RA, Kuhn L;
XX PI MullenbachGL, Parge HE, Tainer JA;
XX DR WPI; 1992-18367/1.22.
XX
PT Fusion proteins with glycosaminoglycan-binding and
PT superoxide-dismutase activities - reduce tissue damage caused by
PT superoxide radicals' useful in treating autoimmune diseases e.g.
PT rheumatoid arthritis and osteoarthritis
XX
PS Claim 8; Fig 1; 140pp: English.
XX
CC The fusion protein comprising the a glycosaminoglycan binding region
CC and human superoxide dismutase, joined via a linker region was
CC established according to the formula A-(2/3)2-SOD where Z is the peptide
CC -RRHHRRKKRRVEED, the fusion protein is useful for extending
CC the in vivo lifetime of biologically active cpds. such as SOD and
CC for targeting them to specific cell surfaces or substrates. The
CC glycosaminoglycan (GAGs) binding protein is formed into a fusion
CC protein with SOD to increase stability, plasma half-life and ease
CC of purification of SOD. SOD is useful for reduction of tissue damage
CC caused by oxygen radicals and is used in the treatment of autoimmune
CC diseases e.g. rheumatoid and osteo-arthritis.
CC See also AAR24225-35, AAR27933-51.
XX
XX Sequence 205 AA:
XX
Query Match 6.0%; Score 167; DB 13; Length 205;
Best Local Similarity 28.1%; Pred. No. 2e-07;
Matches 35; Conservative 32; Mismatches 71; Indels 38; Gaps 8:
QY 185 KARPETAALHHELEEDKTEHYAHCDVRSNTQPKALHH--VHGTDIFKQ-VGYGDLEY 241
DB 10 KRRVEDLMMHHRERKKRVE---DLMAKTAVALVLGGDPVGGILTFQKESNGEVKK 65
QY 242 SYHLEGFNVSDDHKDHLADVOIYANGDLTSGCDNLGATYD-----HEDYSELGDLG 294
DB 66 WGSKTIGLT-----ECLHGPHFHEREDNTAGTSGAPFNPPLSKRHGAPDEERHVGDD 119
QY 295 DIHDDHGVVNESHRYSWVINFQDSDVGLGRSTAIHQK-DLHLK-----SAKTA 341
DB 120 NVYADKKGADVADVIEDSVLISLSDSHLIGRTLVYHEKADLGKGGNEBSTRKGNAGSRLL 179
QY 342 CCVYGRQS-----RP 352
DB 180 CGVYIAQGFGRRHHP 195

```

Search completed: November 29, 2002, 12:32:28
Job time : 74 secs

THIS PAGE BLANK (USPTO)

101,000,000

THE

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2002, 12:32:35 ; Search time 26 Seconds
(without alignments)
562.430 Million cell updates/sec

Title: US-09-868-760-7

Sequence: 1 DGECCNDGKNDKHHDDHD.....MOGHEVSERVACVIGRA 497

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/1/iaa/5A.COMB.pep:*
- 3: /cgn2_6/ptodata/1/iaa/5A.COMB.pep:*
- 4: /cgn2_6/ptodata/1/iaa/5A.COMB.pep:*
- 5: /cgn2_6/ptodata/1/iaa/5A.COMB.pep:*
- 6: /cgn2_6/ptodata/1/iaa/5A.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	173.5	6.2	244	US-08-679-493A-188	Sequence 188, App
2	168	6.0	150	US-08-722-050-9	Sequence 9, App1
3	167	6.0	154	US-08-679-493A-211	Sequence 211, App
4	166	5.9	151	US-08-722-050-10	Sequence 10, App1
5	165.5	5.9	152	US-08-722-050-12	Sequence 12, App1
6	164.5	5.9	153	US-08-679-493A-207	Sequence 207, App
7	164	5.9	151	US-08-679-493A-191	Sequence 191, App
8	163.5	5.9	153	US-08-679-493A-201	Sequence 201, App
9	161.5	5.8	153	US-08-679-493A-202	Sequence 202, App
10	160.5	5.7	152	US-08-679-493A-209	Sequence 209, App
11	159	5.7	166	US-08-679-493A-192	Sequence 192, App
12	158	5.7	151	US-08-679-493A-193	Sequence 193, App
13	158	5.7	152	US-08-679-493A-203	Sequence 203, App
14	157.5	5.6	168	US-08-668-381A-5	Sequence 5, App1
15	156.5	5.6	154	US-08-668-381A-5	Sequence 5, App1
16	156.5	5.6	1021	US-07-910-760-12	Sequence 12, App1
17	156.5	5.6	1021	US-08-440-519-12	Sequence 12, App1
18	156.5	5.6	1021	US-08-440-519-12	Sequence 12, App1
19	156.5	5.6	1021	US-08-440-519-12	Sequence 12, App1
20	155.5	5.6	152	US-08-679-493A-204	Sequence 204, App
21	155.5	5.6	153	US-08-679-493A-204	Sequence 204, App
22	155.5	5.6	153	US-08-679-493A-204	Sequence 204, App
23	155.5	5.6	154	US-09-126-109-4	Sequence 4, App1
24	153.5	5.5	152	US-08-722-050-7	Sequence 7, App1
25	152.5	5.5	152	US-09-291-562-2	Sequence 2, App1
26	152.5	5.5	153	US-08-722-050-3	Sequence 3, App1
27	152	5.4	841	US-08-350-884-86	Sequence 86, App1

28	152	5.4	841	US-08-709-173-86	Sequence 86, App1
29	152	5.4	841	US-08-709-177-86	Sequence 86, App1
30	151	5.4	201	US-08-679-493A-190	Sequence 190, App
31	150	5.4	151	US-08-722-050-8	Sequence 8, App1
32	150	5.4	152	US-08-679-493A-208	Sequence 208, App
33	150	5.4	217	US-08-679-493A-196	Sequence 196, App
34	148	5.3	151	US-08-679-493A-194	Sequence 194, App
35	148	5.3	222	US-08-556-965-2	Sequence 2, App1
36	148	5.3	222	US-08-556-965-2	Sequence 2, App1
37	148	5.3	240	US-08-023-980B-45	Sequence 45, App1
38	148	5.3	240	US-08-486-953A-53	Sequence 53, App1
39	148	5.3	240	US-08-679-493A-186	Sequence 186, App
40	148	5.3	240	US-08-679-493A-186	Sequence 186, App
41	147.5	5.3	309	US-08-101-406-2	Sequence 101, App
42	147	5.3	151	US-08-722-050-3	Sequence 3, App1
43	146.5	5.2	202	US-08-679-493A-197	Sequence 197, App
44	145	5.2	151	US-08-679-493A-193	Sequence 193, App
45	145	5.2	152	US-08-679-493A-206	Sequence 206, App

ALIGNMENTS

```
RESULT 1
US-08-679-493A-188
; Sequence 188, Application US/08679493A
; Patent No. 6303295
; GENERAL INFORMATION:
; APPLICANT: Taylor, Ethan W.
; TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
; FILE REFERENCE: 55-95
; CURRENT APPLICATION NUMBER: US/08/679,493A
; PRIOR FILING DATE: 1996-07-12
; PRIOR APPLICATION NUMBER: 60/001203
; PRIOR FILING DATE: 1995-07-14
; PRIOR APPLICATION NUMBER: 60/003,112
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 188
; LENGTH: 244
; TYPE: PRT
; ORGANISM: rat
US-08-679-493A-188

Query Match      6.2%, Score 173.5, DB:4, Length 244:
Best Local Similarity 29.7%, Pred. No. 1e-08:
Matches 54; Conservative 28; Mismatches 61; Indels 39; Gaps 10;

QY 194 LHHLEEDKTEHYACDVRSNTQPKAL---HHVHGTFDFOVGYGD-LEVSYLEGF 248
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 55 LGRFADARREHNAVCYV-----OPSAFLPPDQPIGTGLVFLQLPSSRLTASFNLGEF 109

QY 249 NVSDHDKHLDHVOIYANGDLTSGCDNLGAKYDPHEHYHSELGLDLDIHDDHG--VYNE 306
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 110 PABQNTSNA--IHVHEFGDLSGCESTGYHNP-----LGVPHQHPGDFGNFYVARD 160

QY 307 ----SHRYEW--INIGDSSVIGRSIAIHQ-RDHLKSA-----KIAACYIGRG 348
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 161 GRMKRMKLAISLAGPSHISILGRVYVHAGEDDLGKGNASVQNGNARRLACCYVGS 220

QY 349 QS 350
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 221 NS 222

RESULT 2
US-08-722-050-9
; Sequence 9, Application US/08722050
; Patent No. 58711729
; GENERAL INFORMATION:
; APPLICANT: YU, GUO-LIANG
; APPLICANT: ROSEN, CRAIG A.
```

```

APPLICANT: FRASER, CLAIRE M.
APPLICANT: COCAVNE, JEANNINE D.
TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNER, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/722,050
FILING DATE: 23-JAN-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/725,757
FILING DATE: 11-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.1020001/EKS/AJK
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 150 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-722-050-9

Query Match 6.0%; Score 168; DB 2; Length 150.
Best Local Similarity 31.9%; Pred. No. 1.7e-08;
Matches 46; Conservative 22; Mismatches 46; Indels 30; Gaps 5.

QY 225 VHGTFIDKQVGYDDLEVSHTLEGFVNSDDKHDLHDVQIYANGSLTSGCNLGAKTDP-- 282
Db 12 VKGTFISQGGDDPTVTYSISGL-----PGLGLGFVHNLSTTNGCKSTGPHNPVG 65
QY 283 -----HEDYHSELGLDGLDHDHDDHGVYVNSHRYSWINIFGDDSVLGRSIAIH-QRDHL 334
Db 66 KKHGAPDEDEHRA--GDLGLNVTAGSDGVVNNITDSQIPLAGPHSIIIGRAVVVHADPDL 123
QY 335 HK-----SAKIACVIG 346
Db 124 GKGGHLSKSTGNAGRGVACGIG 147

RESULT 3
US-08-679-493A-211
Sequence 211, Application US/08679493A
Patent No. 6303295
GENERAL INFORMATION:
APPLICANT: Taylor, Ethan W.
TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
FILE REFERENCE: 55-95
CURRENT APPLICATION NUMBER: US/08/679,493A
PRIOR FILING DATE: 1996-07-12
PRIOR APPLICATION NUMBER: 60/001203
PRIOR FILING DATE: 1995-07-14
PRIOR APPLICATION NUMBER: 60/003,112
PRIOR FILING DATE: 1995-09-01
NUMBER OF SEQ ID NOS: 216
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 211

```

```

?      LENGTH: 154
?      TYPE: PRT
?      ORGANISM: Saccharomyces cerevisiae
US-08-679-493A-211

Query Match                               6.0%; Score 167; DB 4; Length 154;
Best Local Similarity                     30.1%; Pred. No. 2,2e+08;
Matches   43; Conservative    26; Mismatches 48; Indels   26; Gaps     5;

OY  225 VHGTFDKVGVGD-LEVSYLEGPNVSDHKDLHVOIYANGDILTSGCDNGAKYDPH 283
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  14 VSQVFEEQASESEPTTVSYELAGSPRAENGFIHF-----GDATNGCVSAGPIENF 68
OY  284 EDVHS-----ELDDIDGHDDGGVYNESHRSWNTINFGDSYGLRSTAIHQ-RDHLH 335
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  69 KKHGAPTEBVRHVGMGVKTDEGNKAKGSFKDSLKLIGPISYGRSVYIHAGDDLG 128
OY  336 K-----SAKIACVIYG 346
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  129 KGDTESLKTGNAGPRPAGVIG 151

RESULT 4
US-08-722-050-10
Sequence 10, Application US/08722050
Patent No. 5871729
GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
APPLICANT: ROSEN, CRAIG A.
APPLICANT: FRASER, CLAIRE M.
APPLICANT: GOCAYNE, JEANNINE D.
TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNER, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/722,050
APPLICATION NUMBER: US/08/722,050
FILING DATE: 23-JAN-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/725,757
FILING DATE: 11-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488, 1020001/EKS/AJK
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-722-050-10

Query Match                               5.9%; Score 166; DB 2; Length 151;
Best Local Similarity                     30.3%; Pred. No. 2,6e+08;
Matches   43; Conservative    24; Mismatches 49; Indels   26; Gaps     5;

OY  225 VHGTFDKVGVGDLEVSYLEGPNVSDHKDLHVDIYANGDILTSGCNCNLAKKYDPH 284

```

```

Db 13 VQGTIFEFKAGCEPMLSGQITGL-TEGQHGPHYOX-----GDNTOGCTAGPHFNPHS 66
QY 285 DYHS-----ELGDLGDIHDDHGCVNESHRYSMWIFGDDSVLGRSTAIHQ-RDHLK 336
Db 67 KKHGCPADEERHVGDLGNVTAGKDCVANVSTEDRVISLSEHSIIGRTVYVHEKODDLCK 126
QY 337 -----SAKIACCVIG 346
Db 127 GGNESSTKTGNAGSRACGVIG 148

RESULT 5
US-08-722-050-12
; Sequence 12, Application US/08722050
; Patent No. 5871729
; GENERAL INFORMATION:
; APPLICANT: YU, GUO-LIANG
; APPLICANT: ROSEN, CRAIG A.
; APPLICANT: PRASER, CLAIRE M.
; APPLICANT: GOCAYNE, SPERNINE D.
; TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/722,050
; FILING DATE: 23-JAN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/725,757
; FILING DATE: 11-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.1020001/EKS/AJK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2540
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 152 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-722-050-12

Query Match 5.9%; Score 165.5; DB 2; Length 152;
Best Local Similarity 30.8%; Pred. NO. 3e-08;
Matches 44; Conservative 25; Mismatches 47; Indels 27; Gaps 6;
QY 225 VHGTFDKOVGCD-LEVSYLEGFVSDHDKDLHDVOIYANGDLTSGCDNLGAKYDPH 283
Db 13 VSCVVFEDASESEPTVYSYELAG-NSPNAERFHIHF-----GDAVNGCVSAGPHFNP 66
QY 284 EDHS-----ELGDLGDIHDDHGCVNESHRYSMWIFGDDSVLGRSTAIHQ-RDHLK 335
Db 67 KKHGCPADEERHVGDLGNVTAGKDCVANVSTEDRVISLSEHSIIGRTVYVHEKODDLCK 126
QY 336 K-----SAKIACCVIG 346
Db 127 GGNESSTKTGNAGSRACGVIG 148

```

```

RESULT 6
US-08-679-493A-207
; Sequence 207, Application US/08679493A
; Patent No. 6303295
; GENERAL INFORMATION:
; APPLICANT: Taylor, Ethan W.
; TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
; FILE REFERENCE: 55-95
; CURRENT APPLICATION NUMBER: 1996-07-12
; PRIOR APPLICATION NUMBER: 60/001203
; PRIOR FILING DATE: 1995-07-14
; PRIOR APPLICATION NUMBER: 60/003,112
; PRIOR FILING DATE: 1995-09-01
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 207
; LENGTH: 153
; TYPE: PRT
; ORGANISM: horse
; US-08-679-493A-207

Query Match 5.9%; Score 164.5; DB 4; Length 153;
Best Local Similarity 31.0%; Pred. NO. 3.7e-08;
Matches 45; Conservative 26; Mismatches 43; Indels 31; Gaps 6;
QY 225 VHGTFDKOVGYDLEVSYLEGF---NVSDHDKDLHDVOIYANGDLTSGCDNLGAKYD 281
Db 14 VHGVIHFEGQDEGGPVV---LKGFTGLTKGDHGFVHERF-----GDNTOGCTAGAHFN 65
QY 282 P-----HEDYSELGDLGDIHDDHGCVNESHRYSMWIFGDDSVLGRSTAIHQ-RDHLK 333
Db 66 PLKKGKGRDEERHVGDLGNVTADENKADVDKDSVLSKHSIIGRTVYVHEKODD 125
QY 334 LHK-----SAKIACCVIG 346
Db 126 LKGNESTKTGNAGSRACGVIG 150

RESULT 7
US-08-679-493A-191
; Sequence 191, Application US/08679493A
; Patent No. 6303295
; GENERAL INFORMATION:
; APPLICANT: Taylor, Ethan W.
; TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
; FILE REFERENCE: 55-95
; CURRENT APPLICATION NUMBER: 1996-07-12
; PRIOR APPLICATION NUMBER: 60/001203
; PRIOR FILING DATE: 1995-07-14
; PRIOR APPLICATION NUMBER: 60/003,112
; PRIOR FILING DATE: 1995-09-01
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 191
; LENGTH: 151
; TYPE: PRT
; ORGANISM: corn
; US-08-679-493A-191

Query Match 5.9%; Score 164; DB 4; Length 151;
Best Local Similarity 31.7%; Pred. NO. 4.1e-08;
Matches 45; Conservative 21; Mismatches 50; Indels 26; Gaps 4;
QY 225 VHGTFDKOVGYDLEVSYLEGFVSDHDKDLHDVOIYANGDLTSGCDNLGAKYDP-- 282
Db 13 VKGTIFETQEGDPTTVYSYSGL-----KPGDHGFVHALGDTTNGCMSTGPHYNP 66
QY 283 -----HEDYSELGDLGDIHDDHGCVNESHRYSMWIFGDDSVLGRSTAIHQ-RDHLK 336

```

Mon Dec 2 12:20:01 2002

us-09-868-760-7.ra1

Page 4

Db 67 KENGAPEDENRHHGDLGNTAGADGVANINVTDSQIPLTGPNSTIGRAVYVHHADPDOLK 126
OY 337 -----SAKIACCVIG 346
Db 127 GGHLEKSTGNAGRVACITIG 148

RESULT 8
US-08-679-493A-201
Sequence 201, Application US/08679493A
Patent No. 6303295
GENERAL INFORMATION:
APPLICANT: TAYLOR, Ethan W.
TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
FILE REFERENCE: 55-95
CURRENT APPLICATION NUMBER: US/08/679,493A
PRIORITY FILING DATE: 1996-07-12
PRIOR APPLICATION NUMBER: 60/001203
PRIORITY FILING DATE: 1995-07-14
PRIOR APPLICATION NUMBER: 60/003,112
PRIORITY FILING DATE: 1995-09-01
NUMBER OF SEQ ID NOS: 216
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 201
LENGTH: 153
TYPE: PRT
ORGANISM: Mus sp.
US-08-679-493A-201

Query Match 5.9%; Score 163.5; DB 4; Length 153;
Best Local Similarity 30.1%; Pred. No. 4,6e-08;
Matches 44; Conservative 25; Mismatches 50; Indels 27; Gaps 6;
OY 225 VHGTFDKQYGYGD-LEVSYHLEGFNVSDHDKHLDHVOIYANGDLTSGCDNLGAKYDPH 283
Db 14 VOGHIEQKASGEPRVLTSGQITGL-TEGKHGFVHGY-----GDNTOGCTTAGPHFNPH 67
OY 284 EDYHS-----ELGDLGDIHDDHGVNESHRYSWINIFGDDSVLGRSIAIHOR-DHLK 335
Db 68 SKKHGPADEERHVGDLGNVATKDCADVANSIEDSVISLGSBHSIIGRTLVVHKKADLKG 127
OY 336 K-----SAKIACCVIGRQ 349
Db 128 KGNNESTKTGNAGSRILACGVIGIAQ 153

RESULT 9
US-08-679-493A-202
Sequence 202, Application US/08679493A
Patent No. 6303295
GENERAL INFORMATION:
APPLICANT: TAYLOR, Ethan W.
TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
FILE REFERENCE: 55-95
CURRENT APPLICATION NUMBER: US/08/679,493A
PRIORITY FILING DATE: 1996-07-12
PRIOR APPLICATION NUMBER: 60/001203
PRIORITY FILING DATE: 1995-07-14
PRIOR APPLICATION NUMBER: 60/003,112
PRIORITY FILING DATE: 1995-09-01
NUMBER OF SEQ ID NOS: 216
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 202
LENGTH: 153
TYPE: PRT
ORGANISM: rat
US-08-679-493A-202

Query Match 5.8%; Score 161.5; DB 4; Length 153;
Best Local Similarity 30.1%; Pred. No. 7,2e-08;
Matches 44; Conservative 24; Mismatches 51; Indels 27; Gaps 6;
OY 225 VHGTFDKQYGYGD-LEVSYHLEGFNVSDHDKHLDHVOIYANGDLTSGCDNLGAKYDPH 283

Db 14 VOGHIEQKASGEPRVYVSGQITGL-TEGKHGFVHGY-----GDNTOGCTTAGPHFNPH 67
OY 284 EDYHS-----ELGDLGDIHDDHGVNESHRYSWINIFGDDSVLGRSIAIHOR-DHLK 335
Db 68 SKKHGPADEERHVGDLGNVATKDCADVANSIEDSVISLGSBHSIIGRTLVVHKKADLKG 127
OY 336 K-----SAKIACCVIGRQ 349
Db 128 KGNNESTKTGNAGSRILACGVIGIAQ 153

RESULT 10
5171680-3
Patent No. 5171680
APPLICANT: MULLENBACH, GUY T.; HALLEWELL, ROBERT A.; VALEZUELA, PABLO
TITLE OF INVENTION: SUPEROXIDE DISMUTASE ANALOGS HAVING NOVEL BINDING PROPERTIES
NUMBER OF SEQUENCES: 15
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/561,442
FILING DATE: 01-AUG-1990
SEQ ID NO: 3
LENGTH: 152
5171680-3

Query Match 5.7%; Score 160.5; DB 6; Length 152;
Best Local Similarity 29.7%; Pred. No. 8,9e-08;
Matches 43; Conservative 30; Mismatches 45; Indels 27; Gaps 5;
OY 225 VHGTFDKQYGYDLEVSYHLEGFNVSDHDKHLDHVOIYANGDLTSGCDNLGAKYDPH 282
Db 15 VOGHIEQKASGEPRVLTSGQITGL-TEGKHGFVHGY-----GDLHGFVHGFEGDNTAGCTAGPHFNPH 67
OY 283 -----HEDYSELGDLGDIHDDHGVNESHRYSWINIFGDDSVLGRSIAIHOR-DHLK 336
Db 68 RKHGPDEERHVGDLGNVATKDCADVANSIEDSVISLGSBHSIIGRTLVVHKKADLKG 127
OY 337 -----SAKIACCVIGRQ 349
Db 128 KGNNESTKTGNAGSRILACGVIGIAQ 152

RESULT 11
US-08-679-493A-209
Sequence 209, Application US/08679493A
Patent No. 6303295
GENERAL INFORMATION:
APPLICANT: TAYLOR, Ethan W.
TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
FILE REFERENCE: 55-95
CURRENT APPLICATION NUMBER: US/08/679,493A
PRIORITY FILING DATE: 1996-07-12
PRIOR APPLICATION NUMBER: 60/001203
PRIORITY FILING DATE: 1995-07-14
PRIOR APPLICATION NUMBER: 60/003,112
PRIORITY FILING DATE: 1995-09-01
NUMBER OF SEQ ID NOS: 216
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 209
LENGTH: 166
TYPE: PRT
ORGANISM: loggerhead
US-08-679-493A-209

Query Match 5.7%; Score 159; DB 4; Length 166;
Best Local Similarity 28.9%; Pred. No. 1,4e-07;
Matches 41; Conservative 29; Mismatches 46; Indels 26; Gaps 5;
OY 225 VHGTFDKQYGYDLEVSYHLEGFNVSDHDKHLDHVOIYANGDLTSGCDNLGAKYDPH 282
Db 27 VKGTFIEQKASGEPRVLTSSITGL-TEGKHGFVHGF-----GDNTOGCTTAGPHFNPH 80

```

QY   283 -----HEVYHSELDLDDIHDDHGVMNESHRYSWINIFGDDSVLRSTAIHQ--DHLTK 336
      :|::||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    81 KKHGGPQDNERNHVDDGLGNVTANKEGVAEFCIKDSLISLTSSQSILRIQVMVAHEKKRDDLGR 140
QY   337 -----SAKICCVYG 346
      ::|||::|::|
Db    141 GGNDSELKTGNAGSRLLACGYVG 162

RESULT 12
US-08-679-493A-192
; Sequence 192, Application US/08679493A
; Patent No. 6303295
; GENERAL INFORMATION:
; APPLICANT: Taylor, Ethan W.
; TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
; FILE REFERENCE: 55-95
; CURRENT APPLICATION NUMBER: US/08/679,493A
; PRIOR FILING DATE: 1996-07-12
; PRIOR APPLICATION NUMBER: 60/001203
; PRIOR FILING DATE: 1995-07-14
; PRIOR APPLICATION NUMBER: 60/003,112
; PRIOR FILING DATE: 1995-09-01
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 192
; LENGTH: 151
; TYPE: PRT
; ORGANISM: rice
US-08-679-493A-192

Query Match          5.7%; Score 158; DB 4; Length 151;
Best Local Similarity 31.0%; Pred. No. 1.5e-07;
Matches 44; Conservative 21; Mismatches 51; Indels 26; Gaps

QY   225 VHGTIDFKOVGCDLEYSYLGEFNVSDDKRDHLHDYOIYANGDLTSGCNGAKYP-- 282
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    13 VKCTIFFSOGEGCPITYSGVSIG-----KPLHGPFVHALGDPTTNCMGSTGPHPNPTG 66
QY   283 -----HEDVHSLDLDIDHDDHGVMNESHRYSWINIFGDDSVLRSTAIH-QDHCLK 336
      :|::||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    67 KHGAAGPDENHNADDLITAGADGAAANNVSDSQILPTLAHSIIIRAVVVAADPDLGR 126
QY   337 -----SAKICCVYG 346
      ::|||::|::|
Db    127 GGHELSTKTGNAGCGVACGIIG 148

RESULT 13
US-08-679-493A-203
; Sequence 203, Application US/08679493A
; Patent No. 6303295
; GENERAL INFORMATION:
; APPLICANT: Taylor, Ethan W.
; TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
; FILE REFERENCE: 55-95
; CURRENT APPLICATION NUMBER: US/08/679,493A
; PRIOR FILING DATE: 1996-07-12
; PRIOR APPLICATION NUMBER: 60/001203
; PRIOR FILING DATE: 1995-07-14
; PRIOR APPLICATION NUMBER: 60/003,112
; PRIOR FILING DATE: 1995-09-01
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 203
; LENGTH: 152
; TYPE: PRT
; ORGANISM: guineapig
US-08-679-493A-203

Query Match          5.7%; Score 158; DB 4; Length 152;
Best Local Similarity 29.7%; Pred. No. 1.5e-07;

```

Matches	43:	Conservative	26:	Mismatches	50:	Indels	26:	Gaps	5
QY	225	VHGCTIDFQVGYDGLSEVSYHLSEFPNVSDDKDHLHDQIVANDDLTSGCDNIGAKYPP--	282						
DB	14	VGGIIHFQDKANGVYVYKGRITSL-VEKGHGFVHER-----CDNQGSTSAGRPFNLS	67						
QY	283	-----HEPYHSELDGDIHDDHGVNSESRRYSWIMIFGDDSVLGRSLAIHOR-DILHK	336						
DB	68	KKHGQRQDERHNVADLVNTYAGADGVANVSIEDSLTSGANSIIIGRIWVHERFDDLGR	127						
QY	337	-----SAKIACYIGRGQ	349						
DB	128	GGNESTYTGNAISRILAGVIGIAQ	152						
RESULT	14								
	US-08-668-381A-5								
	; Sequence 5, Application US/08668381A								
	; Patent No. 5780024								
	; GENERAL INFORMATION:								
	; APPLICANT: Brown, Robert H.								
	; APPLICANT: Fishman, Paul S.								
	; APPLICANT: Francis, Jonathan W.								
	; APPLICANT: Hosier, Betsy A.								
	; TITLE OF INVENTION: SUPEROXIDE DISMUTASE/TETANUS TOXIN								
	; TITLE OF INVENTION: FRAGMENT C HYBRID PROTEIN								
	; NUMBER OF SEQUENCES: 6								
	; CORRESPONDENCE ADDRESS:								
	; ADDRESSEE: Fish & Richardson P.C.								
	; STREET: 225 Franklin Street								
	; CITY: Boston								
	; STATE: MA								
	; COUNTRY: USA								
	; ZIP: 02110-2804								
	; COMPUTER READABLE FORM:								
	; MEDIUM TYPE: Floppy disk								
	; COMPUTER: IBM PC compatible								
	; OPERATING SYSTEM: PC-DOS/MS-DOS								
	; SOFTWARE: Patentin Release #1.0, Version #1.30								
	; CURRENT APPLICATION DATA:								
	; APPLICATION NUMBER: US/08/668,381A								
	; FILING DATE: 21-JUN-1996								
	; CLASSIFICATION: 514								
	; PRIOR APPLICATION DATA:								
	; APPLICATION NUMBER: 60/000,473								
	; FILING DATE: 23-JUN-1995								
	; ATTORNEY/AGENT INFORMATION:								
	; NAME: Clark, Paul T.								
	; REGISTRATION NUMBER: 30,164								
	; REFERENCE/DOCKET NUMBER: 00786/269001								
	; TELECOMMUNICATION INFORMATION:								
	; TELEPHONE: 617/542-5070								
	; TELEFAX: 617/542-8906								
	; TELEX: 200154								
	; INFORMATION FOR SEQ ID NO: 5:								
	; SEQUENCE CHARACTERISTICS:								
	; LENGTH: 618 amino acids								
	; TYPE: amino acid								
	; TOPOLOGY: linear								
	; MOLECULE TYPE: protein								
	; US-08-668-381A-5								
Query Match		5.6%;	Score 157.5;	DB 1:	Length 618;				
Best Local Similarity		27.8%;	Pred. NO. 1.4e-06;						
Matches	58:	Conservative	56:	Mismatches	64:	Indels	55:	Gaps	10:
QY	225	VHGCTIDFQVQ-VGYDLEVSYHLSEGFNVSDDKDHLHDQIVYANGDITSGCDNIGAKYDP-282							
DB	15	VGGIIHFQDKANGSPKAVWGKISGLT-----EGIHGCVHERGDMTACCTACGAGHNPPL	68						
QY	283	-----HEDYHSELGDIHDDHGVNESHRSYWINIFGDDSVLGRSLAIHOR-DHLH	335						
DB	69	SRKHGAPRQDERHNVADLVNTYAGADGVANVSIEDSLTSGANSIIIGRIVLVNERKADLG	128						

Mon Dec 2 12:20:01 2002

us-09-868-760-7.rai

Page 6

QY 336 K-----SAKIACCYTGKQ-----SHEIYHRAK-----CYVRNTE-----S 368
 Db 129 KGGNESITKGNAGSLACGYIGLAGSVSYSTLPFSYSKINLDCWV-DNEEDIVILKKS 187
 QY 369 TGLAHNVGSSTFEQTPGSGTHMTADLAKGFVVS 401
 Db 188 TILMLDINDI-----ISDISGFVSS 208

```

RESULT 15
5290690-5
: Patent No. 5290690
: APPLICANT: MRAEET, NADIR;LASTERS, IGNACE;STANSENS, PATRICK
: MATTHYSSENS, GASTON;WODAN, SHOSHAN;OON, WILHELMUS J.
: TITLE OF INVENTION: METHODS AND MEANS FOR CONTROLLING THE
: STABILITY OF PROTEINS
: NUMBER OF SEQUENCES: 22
: CURRENT APPLICATION DATA
: APPLICATION NUMBER: 08/
: FILING DATE: 25-AUG-1969
: SEQ ID NO:5:
: LENGTH: 154
5290690-5

```

[illegible]

Search completed: November 29, 2002, 12:36:52
Job time : 29 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2002, 12:33:20 ; Search time 21 seconds
(without alignments)
376,873 Million cell updates/sec

Title: US-09-868-760-7

Perfect score: 2794
Sequence: 1 DGEQNDGQKKDDHDDHND.....MOGHVESERVACVIGRA 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 102317 seqs, 15924203 residues

Total number of hits satisfying chosen parameters: 102317

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

Published_Applications_AA:*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep:*

2: /cgn2_6/ptodata/1/pubpaa/PCOT_NEW_PUB pep:*

3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep:*

4: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep:*

5: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep:*

6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep:*

7: /cgn2_6/ptodata/1/pubpaa/PCOTUS_PUBCOMB pep:*

8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep:*

9: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep:*

10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep:*

11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep:*

12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep:*

13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep:*

14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	168	6.0	150	US-09-883-985-9	Sequence 9, App1
2	166	5.9	151	US-09-883-985-10	Sequence 10, App1
3	165.5	5.9	152	US-09-883-985-12	Sequence 12, App1
4	160	5.7	1099	US-09-881-654-4	Sequence 4, App1
5	155.5	5.5	154	US-09-804-987-6	Sequence 6, App1
6	153.5	5.5	152	US-09-883-985-3	Sequence 7, App1
7	152.5	5.5	153	US-09-883-985-7	Sequence 3, App1
8	150	5.4	151	US-09-883-985-8	Sequence 8, App1
9	147	5.3	151	US-09-883-985-11	Sequence 11, App1
10	142.5	5.1	150	US-09-883-985-11	Sequence 2, App1
11	138	4.9	150	US-09-883-985-4	Sequence 4, App1
12	137	4.9	151	US-09-883-985-6	Sequence 6, App1
13	136.5	4.9	151	US-09-883-985-4	Sequence 2, App1
14	134	4.8	288	US-09-964-899-21	Sequence 21, App1
15	134	4.8	310	US-09-925-297-795	Sequence 795, App
16	118	4.2	292	US-09-864-761-37944	Sequence 37944, A
17	115	4.1	631	US-09-789-561-86	Sequence 86, App1
18	112	4.0	725	US-09-994-485-2	Sequence 2, App1
19	110.5	4.0	409	US-09-813-329-6	Sequence 6, App1

ALIGNMENTS

20	109.5	3.9	544	10	US-09-864-761-48016	Sequence 48016, A
21	107.5	3.8	488	10	US-09-840-787-17	Sequence 17, App1
22	107.5	3.6	842	10	US-09-798-831-8	Sequence 8, App1
23	101.5	3.5	409	10	US-09-813-329-2	Sequence 2, App1
24	97	3.5	724	10	US-09-994-485-4	Sequence 4, App1
25	96	3.4	82	10	US-09-864-761-33313	Sequence 33313, A
26	95.5	3.4	2785	10	US-09-801-574-8	Sequence 8, App1
27	95	3.4	406	10	US-09-813-329-4	Sequence 4, App1
28	94	3.4	584	10	US-09-815-245-5013	Sequence 5013, App
29	94	3.4	589	10	US-09-815-245-10878	Sequence 10878, A
30	91.5	3.3	1838	10	US-09-879-248-8	Sequence 8, App1
31	91	3.3	391	10	US-09-764-870-338	Sequence 328, App
32	90.5	3.2	835	10	US-09-947-199-8	Sequence 8, App1
33	90	3.2	477	10	US-09-966-147-4	Sequence 4, App1
34	90	3.2	842	10	US-09-966-147-2	Sequence 2, App1
35	90	3.2	847	10	US-09-920-859A5	Sequence 5, App1
36	89.5	3.2	835	10	US-09-947-199-2	Sequence 2, App1
37	89.5	3.2	216	10	US-09-764-864-1451	Sequence 1451, App
38	88.5	3.2	1435	10	US-10-153-273-4	Sequence 4, App1
39	88	3.1	222	10	US-09-759-145-479	Sequence 479, App
40	88	3.1	222	10	US-09-780-665-479	Sequence 479, App
41	88	3.1	222	10	US-09-822-827-479	Sequence 479, App
42	88	3.1	243	10	US-09-864-761-43091	Sequence 43091, A
43	87.5	3.1	156	10	US-09-864-761-41679	Sequence 41679, A
44	87.5	3.1	677	10	US-09-737-178-144	Sequence 144, App
45	87.5	3.1	1143	10	US-09-924-154-14	Sequence 14, App1

RESULT 1
US-09-883-985-9
Sequence 9, Application US/09883985
Patent No. US20020081288A1
GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
ROSEN, CLAIRE A.
GOCAYNE, JEANNINE D.
TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESSES:
ADDRESS: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
SYREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/883,985
FILING DATE: 20-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/203,607
FILING DATE: 02-DEC-1998
APPLICATION NUMBER: US 08/722,050
FILING DATE: 23-JAN-1997
APPLICATION NUMBER: US 08/225,757
FILING DATE: 11-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488, 1020003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:
LENGTH: 152 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-883-985-7

Query Match 5.5%; Score 153.5; DB 10; Length 152;
Best Local Similarity 28.7%; Pred. No. 2e-06;
Matches 41; Conservative 30; Mismatches 45; Indels 27; Gaps 5;
OY 225 VHGTFIDKQ-VGYGDLVSYHLEGFNVSDDKHDLADVOIYANGDLTSGCDNLGAKYDP-282
DB 13 VQGTINEOKESGSPKPVKVGSIKGLT-----EGLHGFHVHEFGDNTAGCTSAQPHENPL-66
OY 283 -----HEDYHSELGLGDIHDDHGVNESHVSWINIFGDDSVLGRSLAIHOR-DHLH-335
DB 67 SRKHGPKDERHVGDIQVNTADKGVADVSTEDSVISLSDGHCIIGRITLVHEKADDLG-126
OY 336 K-----SAKIACCVIG-346
DB 127 KGNNESTKTGNAGSRIGCVIG-149

RESULT 7
US-09-883-985-3
Sequence 3, Application US/09883985
Patent No. US20020081288A1
GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
ROSEN, CRAIG A.
FRASER, CLAIRE M.
GOCAYNE, JEANNINE D.
TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P. L. L. C.
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/883,985
FILING DATE: 20-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/203,607
FILING DATE: 02-DEC-1998
APPLICATION NUMBER: US 08/722,050
FILING DATE: 23-JAN-1997
APPLICATION NUMBER: US 08/225,757
FILING DATE: 11-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.1020003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2540
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 153 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-883-985-3

Query Match 5.5%; Score 152.5; DB 10; Length 153;
Best Local Similarity 29.4%; Pred. No. 2.4e-06;
Matches 46; Conservative 28; Mismatches 52; Indels 35; Gaps 9;
OY 357 RACVVRPNTSEGLHHVSGITP-EQTPGSGTHMTADLKGFNVEDLSHHKGVQLHE-415
DB 2 KACVVM--TGTAG--VKGVKFTQETDNGPVHVFNEGSG-----LAKGKHGFVHE-48
OY 416 WGDNSHGCHSLGRMYGHDAHP---KRPDGLDVI--DSDSHGIVHSTRFPDHLN-V-467
DB 49 FGDPTNGCTAGAHFNPTKQEHGAPEDSIRHYGDIGNVVAGADGNVAVNATDKLISLNGS-108
OY 468 EDLANSVLVIMQ-----GQHEVE-----SERVACCVIGRA-497
DB 109 HSLIGRSMVTHENEDDLGRGHLSKVGTGNAGRIACGVGLA-151

RESULT 8
US-09-883-985-8
Sequence 8, Application US/09883985
Patent No. US20020081288A1
GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
ROSEN, CRAIG A.
FRASER, CLAIRE M.
GOCAYNE, JEANNINE D.
TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P. L. L. C.
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/883,985
FILING DATE: 20-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/203,607
FILING DATE: 02-DEC-1998
APPLICATION NUMBER: US 08/722,050
FILING DATE: 23-JAN-1997
APPLICATION NUMBER: US 08/225,757
FILING DATE: 11-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.1020003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2540
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-883-985-8
Query Match 5.4%; Score 150; DB 10; Length 151;

```

Best Local Similarity 30.3%: Pred. No. 3-9e-06; Indels 26; Gaps 4;
Matches 43; Conservative 21; Mismatches 52;

OY      225 VHGTIDPKOVGYDDLEVSHTLEGFNVSDDHKDHLADVOIYNAGDLTSGCDNLGAKYPD-- 282
| | | | | | | | : : : | | | | | | | | | | | | | | | | | | |
Db      13 VSGITLTQVGVAPTYIVNGNISGL-----KPGILGHFVHALGDNTNGCSTGPHTINPAG 66
| | | | | | | | : : : | | | | | | | | | | | | | | | | | | |
OY      283 -----HEDYISELDIGDIHDHDDGVNDESHRYSINIFEGDSVLGSIAIH-QRPHLAK 336
| | | | | | | | : : : | | | | | | | | | | | | | | | | | | |
Db      67 KKHGAPEDEVRAHQDGIGNTIGVEDGTASTFTTDKOIPITGPOSTIGRAVVHADPDOLAK 126
| | | | | | | | : : : | | | | | | | | | | | | | | | | | | |
OY      337 -----SAKIACCVIG 346
| | | | | | | | : : : | | | | | | | | | | | | | | | | | | |
Db      127 GGHELSKSTGCMAGRICAGIIG 148
| | | | | | | | : : : | | | | | | | | | | | | | | | | | | |

RESULT 9
US-09-883-985-5
; Sequence 5, Application US/09883985
; Patent No. US20020081288A1
GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
ROSEN, CRAIG A.
FRASER, CLAIRE M.
GOCCAYNE, JEANNINE D.
TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/883_985
FILING DATE: 20-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/203,607
FILING DATE: 02-DEC-1998
APPLICATION NUMBER: US 08/722,050
FILING DATE: 23-JAN-1997
APPLICATION NUMBER: US 08/225,757
FILING DATE: 11-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.1020003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO.: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-09-883-985-5

Query Match          5.3%; Score 147; DB 10; Length 151;
Best Local Similarity 30.3%; Pred. No. 7.1e-06;
Matches 43; Conservative 21; Mismatches 52; Indels 26; Gaps 4;

OY      225 VHGTIDPKOVGYDDLEVSHTLEGFNVSDDHKDHLADVOIYNAGDLTSGCDNLGAKYPD-- 283
| | | | | | | | : : : | | | | | | | | | | | | | | | | | | |

```

```

Db      13 VKGIPTFTHGNCATVTIGTSGCL-----RPLGLGHVHALDDNNTNGCMSGSPHEHNPDG   66
Oy      284 -----EDYSELSDGDIDHHDDHCVVWVESHRYSWMINIFGDDSVLGRSLAIH-QRHLLHX   336
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      67 KTHGAPEPANRAHGAGDGGNIITVGGDTGTAFITTDSQIPLSGPNSIVGRAIVAHADPDLDK    126
Oy      337 -----SAKIACCVCYG 346
           ::||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      127 GGHELSTGNAGGRVACG 148

RESULT 10
US-09-883-985-11
; Sequence 11, Application US/09883985
; Patent No. US20020081288A1
; GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
ROSEN, CRAIG A.
FRASER, CLAIRE M.
GOCCAYNE, JEANNINE D.
TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESSES:
ADDRESS: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0., Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/883,985
FILING DATE: 20-Jun-2001
CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/203,607
FILING DATE: 02-DEC-1998
APPLICATION NUMBER: US 08/722,050
FILING DATE: 23-JAN-1997
APPLICATION NUMBER: US 08/225,757
FILING DATE: 11-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.1020003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 150 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SOURCE DESCRIPTION: SBQ ID NO: 11:
US-09-883-985-11

Query Match          5.1%; Score 142.5; DB 10; Length 150;
Best Local Similarity 28.2%; Pred. No. 1.7e-05;
Matches 40; Conservative 25; Mismatches 50; Indels 27; Gaps 5;

Oy      285 DYHS-----ELGDLJDIHDDHGVNESHRYSWNINIFGDDSVLGRSIAIHQR-DHLRK   336
           ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      13 VKGVARE-----DQNVVEGRIBELT-----DCNHGFHIHFEGDNTNGCLSGCPHPHNPNON   66
           ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Oy      67 KNHGSPPKADARHGVDGNAV-TAEGVGAQENFTDPDISLKGERSLTGRTAVVHKODDLK 125
           ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

OY 337 -----SAKJACCVIG 346
Db 126 GGDESLKTNMAGRLACGVIG 147

RESULT 11
US-09-883-985-2
Sequence 2, Application US/09883985
Patent No. US20020081288A1

GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
ROSEN, CRAIG A.
FRASER, CLAIRE M.

TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/883,985
FILING DATE: 20-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/203,607
FILING DATE: 02-DEC-1998
APPLICATION NUMBER: US 08/722,050
FILING DATE: 23-JAN-1997
APPLICATION NUMBER: US 08/225,757
FILING DATE: 11-APR-1994

ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.1020003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2540
TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 255 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-883-985-2

Query Match 4.9%; Score 138; DB 10; Length 255;
Best Local Similarity 33.3%; Pred. No. 8.7e-05;
Matches 36; Conservative 19; Mismatches 33; Indels 20; Gaps 4;
OY 409 HGVOLEHGDMSHGCHSLGRMY-----HGHDADHPKRPDGLDGVYDSDSHGIVHSTRFT 462
Db 108 HGLVHVGCDGLTNNCSGHNFPDASHGPGQSDSDRHR-GDLGNVRADADGRAIFRMD 166
OY 463 DHLVNEDELNARSIVIMO-----GGHEV-----ESERVACCVIGRA 497
Db 167 EOLKWDVIARSLIIDGEGDGLGCGHPLSKITNGSGERLACGIARS 214

RESULT 12
US-09-883-985-4
Sequence 4, Application US/09883985

Patent No. US20020081288A1
GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
ROSEN, CRAIG A.
FRASER, CLAIRE M.

TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/883,985
FILING DATE: 20-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/203,607
FILING DATE: 02-DEC-1998
APPLICATION NUMBER: US 08/722,050
FILING DATE: 23-JAN-1997
APPLICATION NUMBER: US 08/225,757
FILING DATE: 11-APR-1994

ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.1020003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2540
TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 150 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-883-985-4

Query Match 4.9%; Score 137; DB 10; Length 150;
Best Local Similarity 29.5%; Pred. No. 5.2e-05;
Matches 44; Conservative 16; Mismatches 55; Indels 34; Gaps 6;
OY 54 VHGSILOKSGHAGVYLELHVGNTSESDHNDHNLHMLGDMASGDSIGELYNANP 113
Db 13 VOGTTHFEKAGDTV-----VTGSTITGLTEGD--HGFMHOGSDWTGCTGAGPFPPLS 65
OY 114 EKRADP-----GDLGDLVDDRGVYNEVNHAYMLDIG--TAPNTEALIGSMITLO 163
Db 66 KKHGKDEERHVGDLGVATDKMGV-----AYDIVDPLISLGEYSITIGRTMYVNE 118
OY 164 -----GSHTDADTPASRIACVIG 182
Db 119 KPDDLGRGNESESTYTGNAAGRLACGVIG 147

RESULT 13
US-09-883-985-6
Sequence 6, Application US/09883985
Patent No. US20020081288A1
GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
ROSEN, CRAIG A.
FRASER, CLAIRE M.

Mon Dec 2 12:20:03 2002

us-09-868-760-7.rapb

Page 7

```

GOCANE, JEANNINE D.
TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSER: STERNER, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/883,985
FILING DATE: 20-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/203,607
FILING DATE: 02-DEC-1998
APPLICATION NUMBER: US 08/722,050
FILING DATE: 23-JAN-1997
APPLICATION NUMBER: US 08/225,757
FILING DATE: 11-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488,1020003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
STRANDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-883-985-6

Query Match
Best Local Similarity 4.8%; Score 136.5; DB 10; Length 151;
Matches 44; Conservative 21; Mismatches 49; Indels 27; Gaps 6;
QY 227 GTIDKQVGV-DLEVYHLEGNNVSDHKLHDVQIYANGDLTSCDNLGAKYDHPED 285
DB 13 GTVEBOSSSTPKVSGVGL-----AKLGHFVHEFGDVTNCGSSGPHFPGK 66
QY 286 YHS-----ELDIDDIHDDHGVVNSHRYSWINIFGDDSVLGRSIAIH-----QR 331
DB 67 EHGAPVDENHLDGNIERATGDCPTKYNTIDSKITLFGADSIIGRTVVVHADADDLGG 126
QY 332 DR-LHRS-----AKINCVYIG 346
DB 127 GHELSKSTGNAGARICGVIG 147

RESULT 14
US-09-964-899-21
Sequence 21, Application US/09964899
Patent No. US2002017446A1
GENERAL INFORMATION:
APPLICANT: Cohen, Dalia et al.
TITLE OF INVENTION: Identification of Genes Involved in
FIBROSIS
REFERENCE: 4-31612 A
CURRENT APPLICATION NUMBER: US/09/964,899
CURRENT FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: 60/236,893
PRIOR FILING DATE: 2000-09-29
```

```

PRIOR APPLICATION NUMBER: 60/236,309
PRIOR FILING DATE: 2001-06-14
NUMBER OF SEQ ID NOS: 53
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 21
LENGTH: 288
TYPE: PRT
ORGANISM: Homo Sapien
US-09-964-899-21

Query Match
Best Local Similarity 4.8%; Score 134; DB 9; Length 288;
Matches 35; Conservative 19; Mismatches 34; Indels 20; Gaps 4;
QY 409 HGVLHEWGMDSHGCHSLGRMY-----HGHDADHPKRRGDLGVDYIDSHGIVSTRTF 462
DB 141 HGLVHQYGDLTNNCSGCHNFPDGAHSGGPOSDRRH-GDLGNVRADADRATFRMED 199
QY 463 DHLNVEDLNARSLVIMQ-----GGHEV-----ESERVACVIGRA 497
DB 200 EQLKWDVIGRSLIDEGEDDLGRGHPLSKITGNSGERLACGIARS 247

RESULT 15
US-09-925-297-795
Sequence 795, Application US/09925297
Patent No. US20020081659A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA105
CURRENT APPLICATION NUMBER: US/09/925,297
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05989
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
NUMBER OF SEQ ID NOS: 928
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 795
LENGTH: 310
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (72)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-795

Query Match
Best Local Similarity 4.8%; Score 134; DB 10; Length 310;
Matches 35; Conservative 19; Mismatches 34; Indels 20; Gaps 4;
QY 409 HGVLHEWGMDSHGCHSLGRMY-----HGHDADHPKRRGDLGVDYIDSHGIVSTRTF 462
DB 163 HGLVHQYGDLTNNCSGCHNFPDGAHSGGPOSDRRH-GDLGNVRADADRATFRMED 221
QY 463 DHLNVEDLNARSLVIMQ-----GGHEV-----ESERVACVIGRA 497
DB 222 EQLKWDVIGRSLIDEGEDDLGRGHPLSKITGNSGERLACGIARS 269

Search completed: November 29, 2002, 12:37:34
Job time : 24 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2002, 12:31:24 ; Search time 44 Seconds

(without alignments)
1085.882 Million cell updates/sec

Title: US-09-868-760-7

Sequence: 1 DGEQCNDGQNKDDHDDHDD...MGGHEVESERVACVIGRA 497

Scoring table: BLOSUM62

[illegible][illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Listing first 45 s

Database : PIR_73:*

```
1: P111:
2: P112: *
```

```
3: p1r3:*
4: p1r4:*
```

Pred. No. is the number of re

score greater than or equal to 100 and is derived by analysis of

Result	%	Query
Success	99.99	10000
Failure	0.01	10000
Timeout	0.00	10000
Success	99.99	100000
Failure	0.01	100000
Timeout	0.00	100000
Success	99.99	1000000
Failure	0.01	1000000
Timeout	0.00	1000000

No.	Score	Match Length	DB	IL
-----	-------	--------------	----	----

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
No.							
1	213	7.6	351	1	KG20HL		histidine-rich gly
2	204.5	7.3	735	2	T43059		hypothetical prote
3	178	6.4	152	2	UW0074		superoxide dismuta
4	178	6.4	852	2	A340373		histidine-rich cal
5	174.5	6.2	251	2	S52859		superoxide dismuta
6	173.5	6.2	152	2	T06570		superoxide dismuta
7	173.5	6.2	244	2	A49097		superoxide dismuta
8	173	6.2	1840	2	T29091		transferrin - chick
9	168	6.0	151	2	A29077		superoxide dismuta
10	167	6.0	154	1	DSBYC		superoxide dismuta
11	164.5	5.9	154	1	DSHOC2		superoxide dismuta
12	164	5.9	152	2	S07007		superoxide dismuta
13	163	5.8	152	2	S22508		superoxide dismuta
14	163	5.8	152	2	S72235		superoxide dismuta
15	162.5	5.8	154	2	TJ00915		superoxide dismuta
16	161.5	5.8	154	2	TJ09192		superoxide dismuta
17	161	5.8	152	2	S36108		superoxide dismuta
18	159.5	5.7	699	2	A54608		superoxide dismuta
19	159	5.7	166	2	S29782		histidine-rich cal
20	158	5.7	152	1	DSPCRY		superoxide dismuta
21	158	5.7	152	2	S21136		superoxide dismuta
22	157.5	5.6	141	2	S20512		superoxide dismuta
23	157	5.6	153	1	S33162		superoxide dismuta
24	155.5	5.6	153	1	DSHUC2		superoxide dismuta
25	154	5.5	179	4	T43640		superoxide dismuta
26	153.5	5.5	222	1	DSSE8C		superoxide dismuta
27	153	5.5	154	2	S24971		superoxide dismuta
28	153	5.5	162	2	T51131		superoxide dismuta
29	152.5	5.5	153	2	A49241		superoxide dismuta

ALIGNMENTS

30	152	5.4	152	1	DSM27	superoxide dismutase
31	150.5	5.4	201	2	T063229	probable superoxide
32	150	5.4	152	2	S04623	superoxide dismutase
33	150	5.4	152	2	S08350	superoxide dismutase
34	150	5.4	152	2	T09778	superoxide dismutase
35	150	5.4	217	2	S48021	superoxide dismutase
36	149.5	5.4	151	1	S09568	superoxide dismutase
37	148.5	5.3	184	2	S40984	superoxide dismutase
38	148.5	5.3	221	2	JE00984	superoxide dismutase
39	148	5.3	152	2	S40404	superoxide dismutase
40	148	5.3	152	2	T07925	superoxide dismutase
41	148	5.3	156	2	T12304	superoxide dismutase
42	148	5.3	240	1	DSH0EC	superoxide dismutase
43	148	5.3	3119	2	T18414	protein g77 - mal
44	147.5	5.3	153	2	S63436	superoxide dismutase
45	147.5	5.3	309	2	A25942	histidine/alanine-

RESULT 1

B:Histidine-rich glycoprotein precursor - Plasmodium lophurae
 C:Species: Plasmodium lophurae
 C>Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 16-Jul-1999
 C:Accession: J22692
 C:Reviewer: J. V. Feder
 C:Nature 312: 616-620 (1984)
 A:Title: Primary structure and genomic organization of the histidine-rich protein of
 A:Accession: A22692; M01D:85061618; PMID:6095114
 A:Molecule: 1 type: DNA
 A:Residue: 1-351 <NAV>
 C:Cross references: GB:X01469; MID:g9997; PIDN:CA25698.1; PID:g9999
 C:Comment: There are two copies of 16-residue repeats, two copies of 17-residue repeats.
 C:Genetics: 23/3
 A:Initrons: 23/3
 C:Superfamily: plasmodium histidine-rich protein
 C:Keywords: glycoprotein; tandem repeat
 F:1-23/Domain: signal sequence #status predicted <SIG>
 F:24-31/Domain: propeptide #status predicted <PRO>
 F:32-31/Product: histidine-rich glycoprotein #status predicted <MAT>
 F:33-107/Region: 16-residue repeats
 F:34-107,108-122/Region: 17-residue repeats
 F:123-301,312-351/Region: 10-residue repeats
 F:14-38,139-151/Region: 10-residue repeats
 F:15-301,312-351/Region: 10-residue repeats
 F:40/Binding site: carboxylate (asn) (covalent) #status predicted

Query Match	7.6%;	Score 213;	DB 1;	Length 351
Post Local Similarity	10.48;	Score 48-08;		

Matches 83; Conservative 31; Mismatches 163; Indels 150; Gaps 17;

[illegible]

[illegible]

RESULT 2

hypothetical protein y3956b.g9 [imported] *Caeorhabdittis elegans*
C/Species: *Caeorhabdittis elegans*
C/Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C/Accession: T45059
R/Author: R. Almscoough, R. Anderson, K. Baynes, C. Berke, M. Bonfield, J. Burton,
raser, A. Fulton, L. Gardner, A. Green, P. Hawkins, T. Hillier, L. Jier, M. Johns
B.: O'Callaghan, M.; Parsons, J.; Percy, C.; Rifkin, L.; Roopra, A.; Saunders, D.
Nature 368, 32-38, 1994
A/Authors: Showkneen, R.; Sims, M.; Smaldon, N.; Smith, A.; Smith, M.; Sonhammer, E.; S
lock, L.; Wilkinson-Sprat, J.; Wohlman, P.
A/Title: 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*.
A/Reference number: S45331; MUID:94150718; PMID:7906398
A/Accession: T45059
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-735 <WILD>
A/Cross-references: EMBL:AL132896; NID:g6434440; PTDN:CAB60938.1; PID:g6434473
A/Experimental source: Clone y39B6b
C/Genetics:
A/Map position: 3
A/Intons: 18/1; 69/1
A/Note: y39B6b.g9

Query Match	7.38	Score 204.5	DB 2	Length 735
Best Local Similarity	21.08	Pred. No. 4.2e-07		
Matches 102; Conservative	29	Mismatches 178;	Indels 177;	Gaps 23

[illegible]

QY	464	HLNVED	469
Db	706	HDDKEN	711

RESULT 3

superoxide dismutase (EC 1.15.1.1) (Cu-Zn) - soybean
N:Alternate names: Cu/Zn-SOD
C:Species: Glycine max (soybean)
C:Date: 18-Jun-1998 #sequence-revision 10-Jul-1998 #text-change 20-Apr-2000
C:Accession: J00084
R:Arachita, M.; Nong, V.H.; Kadokura, K.; Kimura, K.; Udaoka, K.; Fukazawa, C.
Bloscl, B:otechnol. Biochem. 62, 1018-1021, 1998
A:Title: Molecular cloning and expression patterns of Cu/Zn-superoxide dismutases in
A:Reference number: J00084; MUID:98312043; PMID:9648237
A:Accession: J00084
A:Molecule type: mRNA
A:Residues: 1-152 <RNA>
C:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxy
C:Superfamily: superoxide dismutase (Cu-Zn)
C:Keywords: copper; homodimer; metalloprotein; oxidoreductase; zinc
E:45-47/62,119/Binding site: copper (His) #status predicted
E:56-145/Disulfide bonds: #status predicted
E:62,70,79,82/Binding site: zinc (His, His, His, Asp) #status predicted
E:146/Active site: Arg #status predicted

Query Match	6.48;	Score 178;	DB 2;	Length 152;
-------------	-------	------------	-------	-------------

Matches	47;	Conservative	23;	Mismatches	46;	Indels	26;	Gaps	4;
---------	-----	--------------	-----	------------	-----	--------	-----	------	----

225 VHGTIDFKQVGCDLEVSYHLEGFNVSDDHKDHLHDVQIYANGDLTSGCDNLGAKKYDPH - 283

DB 14 VTGTIEFTQEGNGPTTVTGS LAGL-----KPG L HGFHVHALGDTTNGCLSTGAHFNPN 67

QY 284 -----EDYHSELGDLGDHDDHGVNESHRYSWINIEGDDSVLGRSAIH-QRDHLHK 336

Db 68 NEHGAPEDENRHAGDLGNVNVCGDDGTVSFSITDSQIPTGPSNIGRAVVHADSDDLGK 127

QY 337 -----SAKIACCIVIG 346

2

RESULT 4
A34373
histidine-rich calcium-binding protein precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 08-Jun-1990 #sequence_revision 08-Jun-1990 #text_change 05-Nov-1995
A:Accession: A34373
R:Hotmann, S.L.; Goldstein, J.L.; Orth, K.; Moomaw, C.R.; Slaughter, C.A.; Brown, M.S.
J. Biol. Chem. 264, 18083-18090, 1989
A:Title: Molecular cloning of a histidine-rich Ca(2+)-binding protein of sarcoplasmic
A:Reference number: A34373; MUID:90036684; PMID:2808365
A:Accession: A34373
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-852 <HOE>
A:Cross-references: GB:J05080; NID:9165099; PIDN:AAA1279.1; PID:9165100
C:Keywords: calcium binding

Query Match 6.48; Score 178; DB 2; Length 852;

Matches 126; Conservative 54; Mismatches 176; Indels 212; Gaps 38;

QY 1 DGEQNDGQNKDDHDDH --- HDDHHDDHDDDDDETMHYAQCEMERPNRHMASSLHHVHG 56

```

Db 237 DDDDEGSTDHQAHRHGRGHEEEDEDDDE---GDSTESDRHQA---HRH--- 285

```

0V 57 SIELSOKG-----GAVYLELHV-----GNTSEDHDD-----HHGL 90

```
Db 286 -----RGHREDEDEDEGSDTESDRQAHNRHGREEDDDDDDEGSDTESDRQA 339
OY 91 HLHM-----LGMASAGODSGELVYNAH-----PEKNAHPRGLADLVDDGVVNEVH 137
Db 340 HNRHGREDEDEDEDESDTESDRQAHNRHGREDEDEDEDESDTESDR-----HQA 396
OY 138 HYAM-----LDIDGAPRTALIGHSTIL-----QGSHTDADT-PASR1 176
Db 397 RHNGHREDEDEDEDESDTESDRQAHNRHGREDEDEDEDESDTESDRQA 454
OY 177 ACCVIGHGARPE--TAALHNEEL-----EDKTE-HYAH----- 208
Db 455 R---GHEEEDDDDDDEGENHNRHNRHGREDEDEDEGDDDDDDSTENQHNRQHG 510
OY 209 ---CDVRSNTHOPKALHNHV--HGTIDFKOVGDDLEVSYLEGFNSDDKDLHDVOI 263
Db 511 KEAEVTSDEH-----HHNVPRHG-----HOGHDKDEEE--EGVST-----DHMHQVPR 553
OY 264 YAN---GDLTSGCDNLGAKYPRHEDYSELGDLG-DIHDDHGVV--NESHRYSWINI 315
Db 554 HAHNGGEGEEGDEELTYKAGHNVAHSHRPGHRSRGHAEHGTVEVRHNR----- 607
OY 316 FGDSVIGRSIALIHORDHLKSAKJACCVIGRGSHPETVYRAKCVVRPTTESTGHHNV 375
Db 608 -----MGOTDSARGHNPASSP-----ROGHP-----PBDT-VHNR- 638
OY 376 SGSTFE-----OTRG-----GSTNMTADLKGFNVSDDLSHNRGVDLHENGDM 420
Db 639 KGLAKEVGPESFPAGVAKDGSRYKRGSGSEEBEOKG-----TTH-HLEDEDEDEEG 690
OY 421 HGCHSLGRMTHGHDADHPKRPGLADY 448
Db 691 HG-RSLSQ-----EDQEEEDRGESAKY 712

RESULT 5
552859
superoxide dismutase (EC 1.15.1.1) (Cu-Zn) - mouse
C:Species: Mus musculus (house mouse)
C>Date: 06-Jun-1995 #sequence_revision 21-Jul-1995 #text_change 05-May-2000
C:Accession: I48335; S52859
R:Carlsson, L.M.; Jonsson, J.; Edlund, T.; Marklund, S.L.
Proc. Natl. Acad. Sci. U.S.A. 92, 6264-6268, 1995
A:Title: Mice lacking extracellular superoxide dismutase are more sensitive to hypoxemia
A:Reference number: I48335; MUID:95327627; PMID:7603981
A:Accession: I48335
A:Status: preliminary; translated from GB/EMBL/DDB
A:Molecule type: DNA
A:Residues: 1-251 <RES>
A:Cross-references: EMBL:X84940; NID:9695628; PIDN:CAA5935.1; PID:9695629
C:Genetics:
A:Gene: SOD3
C:Function:
A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxygen
C:Superfamily: superoxide dismutase (Cu-Zn)
C:Keywords: copper; metalloprotein; oxidoreductase; zinc
F:128,130,145/Binding site: copper (His) #status predicted
F:145,153,156,159/Binding site: zinc (His, His, Asp) #status predicted
F:218/Active site: Arg #status predicted

Query Match 6.2%; Score 174.5; DB 2; Length 251;
Best Local Similarity 31.1%; Pred. No. 1.e-05;
Matches 56; Conservative 22; Mismatches 57; Indels 45; Gaps 11;

OY 199 EEDKTEHYACHDVRSNTHOPKAL-----HHNVGTDKOVGYGD-LEVSYLEGFNSDD 253
Db 67 EVAAAEHMAHICRY-----OPSAITLPRDPOQTGLVLFRLQDLGSRLEAVYSLEGPFAEON 121
OY 254 HKD---HLNDVOIYANGDITSGCDNLGAKYPRHEDYSELGDLGDIHDDHG--VYNES- 307
Db 122 ASNRATIHVEF-----GDLSGCSDSTGPHYNPMEVPRHQ-----HPSGFGFVVVRNQ 169
OY 308 ---HRS-WLNIFFGDSVIGRSIALIHQ-RDLHLKSA-----KIACCVIGRGS 350
```

```
Db 170 LMRHRYGLASLAGPNSHISIGRVSVVYHAGEDDLGGKGNASLQNGNRRLACCVGTSSS 229

RESULT 6
106570
superoxide dismutase (EC 1.15.1.1) (Cu-Zn) - garden pea
C:Species: Pisum sativum (garden pea)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: T06570
R:White, D.A.; Zillinskas, B.A.
Plant Physiol. 96, 1391-1392, 1991
A:Title: Nucleotide sequence of a complementary DNA encoding pea cytosolic copper/zinc
A:Reference number: 215766
A:Accession: T06570
A:Status: preliminary; translated from GB/EMBL/DDB
A:Molecule type: mRNA
A:Residues: 1-152 <WHI>
A:Cross-references: EMBL:M63003; NID:9169069; PIDN:AAA3659.1; PID:9169070
C:Function:
A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxyg
C:Superfamily: superoxide dismutase (Cu-Zn)
C:Keywords: copper; metalloprotein; oxidoreductase; zinc
F:45,47,62,119/Binding site: copper (His) #status predicted
F:55-145/disulfide bonds: #status predicted
F:62,70,79,82/Binding site: zinc (His, His, His, Asp) #status predicted
F:112/Active site: Arg #status predicted

Query Match 6.2%; Score 173.5; DB 2; Length 152;
Best Local Similarity 31.4%; Pred. No. 1.e-05;
Matches 49; Conservative 26; Mismatches 48; Indels 33; Gaps 5;

OY 211 VRSNTHOPKALHNHVHGTIDFKOVGYGDLEVSYLEGFNSDDKDLHDVOIYANGDIT 270
Db 7 VLSNSNE-----VSGTIFESQEGNGPTVYGTLAGL-----KPLHGFHIALGDTT 53
OY 271 SGGDNLGAKYDPR-----EDYHSELGDLGDIHDDHGVVNESHRYSWINIGDSDVIG 323
Db 54 NCISITGPHNPNGKRGHAPDEDETHAGDGLCNINVCDDGVFTITDNIPLRTGTSIIG 113
OY 324 RSLAIIH-QBDHLHK-----SAKIKCVIYG 346
Db 114 RAVVYHADDDDLGGKGNHLSKTTGNAGGRVACGII 149

RESULT 7
A49097
superoxide dismutase (EC 1.15.1.1) (Cu-Zn) B precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 05-May-2000
C:Accession: A49097; S34053; S37408
R:Williams, J.; Zwijsen, A.; Siegers, H.; Nicotale, S.; Bettadapura, J.; Raymackers, J.
J. Biol. Chem. 268, 24614-24621, 1993
A:Title: Purification and sequence of rat extracellular superoxide dismutase B secret
A:Reference number: A49097; MUID:94043314; PMID:8227019
A:Accession: A49097
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-244 <WHI>
A:Cross-references: EMBL:Z24721; NID:9404250; PIDN:CAA80849.1; PID:9404251
R:Perry, A.C.F.; Jones, R.; Hall, L.
Biochem. J. 293, 21-25, 1993
A:Title: Isolation and characterization of a rat cDNA clone encoding a secreted super
A:Reference number: S34053; MUID:93319511; PMID:8328962
A:Accession: S34055
A:Molecule type: mRNA
A:Residues: 1-234, 'W', 236, 'W', 238-244 <PER>
A:Cross-references: EMBL:X68041; NID:9394738; PIDN:CAA48177.1; PID:9394739
C:Function:
A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxyg
C:Superfamily: superoxide dismutase (Cu-Zn)
C:Keywords: copper; metalloprotein; oxidoreductase; zinc
F:121,123,138,168/Binding site: copper (His) #status predicted
```


A:Accession: A92281
A:Molecule type: protein
R:Residues: 2-55,'D','57-92','D','94-154 <STE>
R:Ramezani Rad, M.; Kirchraath, L.; Hollenberg, C.P.
Submitted to the Protein Sequence Database, September 1995
A:Reference number: S57111
A:Accession: S57125
A:Molecule type: DNA
A:Residues: 1-154 <RAW>
A:Cross-references: EMBL:49604; NID:g1015811; PIDN:CAAB634.1; PTD:g1015812; GSPDB:GNOC
R:Horiuchi, S
Submitted to the Protein Sequence Database, March 1992
A:Reference number: A40093
A:Accession: A40093
A:Molecule type: protein
R:Residues: 2-7,'KL','10-42','L','44-55','D','57-88','L','90-91','AD','94-100','Q','102','K','104-107'
C:Genetics:
A:Gene: SOD; MIPS:YJR104C
A:Cross-references: SGD:S003865; MIPS:YJR104C
A:Map position: 10R
C:Function:
A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxygen
C:Superfamily: superoxide dismutase (Cu-Zn)
C:Keywords: copper; metalloprotein; oxidoreductase; zinc
F:2-154/Product: superoxide dismutase (Cu-Zn) #status experimental <MAY>
F:47,49,64,121/Binding site: copper (His) #status predicted
F:58-147/Dissulfide bonds: #status experimental
F:64,72,81,84/Binding site: zinc (His, His, Asp) #status predicted
F:144/Active site: Arg #status predicted

Query Match 6.0% Score 167: DB 1: Length 154;
Best local similarity 30.1%; Pred. No. 3e-05;
Matches 43; Conservative 26; Mismatches 48; Indels 26; Gaps 5;

Oy 225 VHGIDPKVGYGD-LEVSYLIEGFNVSDHKDHMDVQIYANDLSGGDNIGAKYDDH 283
| | : | : | | | | | : | : | : | : | : | : | : |
Db 14 VSGVVRKEGASESEPTTYSYEILAGSNPNAERGFHNHF-----GDATNGCVSAGPRFNFP 68
Oy 284 EDVYS-----ELGLGDIHDDHGVNESHRSRWNIIFEDSDVLGRSLAIHQ-RDHLH 335
| | : | : | | | | | : | : | : | : | : | : | : |
Db 69 KTHIGATDEVRIVGDMGVNTDENGVAKGSFKDSLILITPSVGRSVVIHAGDDLG 128
Oy 336 K-----SAKIACCVYG 346
| | : | : | | | | | : | : | : | : | : | : | : |
Db 129 KGDTRESLKGTGNAGPRACGVIG 151

RESULT 11
DSHOXZ
superoxide dismutase (EC 1.15.1.1) (Cu-Zn) 1 [validated] - horse
C:Species: Equus caballus (domestic horse)
C:Date: 02-Apr-1982 #sequence-revision 09-Aug-1997 #text-change 20-Apr-2000
C:Accession: JCS5215; A00515
R:de la Rúa-Domech, R.; Wiedmann, M.; Mohamed, H.O.; Cummins, J.F.; Bae
Reine 178, 83-88, 1996
A>Title: Equine motor neuron disease is not linked to Cu/Zn superoxide dismutase mutation
A:Reference number: JCS5215; MUID:97080551; PMID:8921896
A:Accession: JCS5215
A:Molecule type: mRNA
A>Status: preliminary
A:Cross-references: GB:U38956; NID:g1228115; PIDN:AAC48662.1; PTD:g1228116
R:Experimental source: leukocyte
R:Bierich, K.; Ammer, D
J Biol Chem 256, 11545-11551, 1981
A>Title: Amino acid sequence of copper-zinc superoxide dismutase from horse liver.
A:Reference number: A00515; MUID:82052979; PMID:7286516
A:Accession: A00515
A:Molecule type: protein
A:Residues: 2-154 <LER>
C:Genetics:
A:Gene: sod1
C:Function:

```

A>Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxygen
C:Superfamily: superoxide dismutase [Cu-Zn]
C:Keywords: acetylated amino end; copper; metalloprotein; oxidoreductase; zinc
F:2.Modified site: acetylated amino end (Ala) (in mature form) #status experimental
F:47-49,64,121/Binding site: copper (His) #status predicted
F:58-147/Dissulfide bonds: #status predicted
F:64,72,81,84/Binding site: zinc (His, His, Asp) #status predicted
F:144/Active site: Arg #status predicted

Query Match          5.9%; Score 164.5; DB 1; Length 154;
Best Local Similarity 31.0%; Pred.No. 4,6e-05;
Matches 45; Conservative 26; Mismatches 43; Indels 31; Gaps 6;

Db      15 VHGVIHFEQDOEGRPV---LKGFTGLTKDGHGFHVHF-----GDNTOGCCTTAAMHN 66
|||:::||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
OY     225 VHGTIDFKGVGYGDLEVSYLEGF--NVSDHKDHLDHVOIYANGDLTSGCDNLGAKYD 281
|||:::||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

Db      P-----HEDYSIELGLDGIHDHDDHGVAVESRYSNINIFGDSYLKSAIAHOR-DH 333
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
OY     282 P-----HEDYSIELGLDGIHDHDDHGVAVESRYSNINIFGDSYLKSAIAHOR-DH 333
||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      PLAKKHGGPDEERHHVGDIIGNVTADENCKADVDKMKSIVISLSGKNHSITIGRTVMVHEKKOD 126
|||:::||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

OY     334 LHR-----SARJACCVIG 346
|||:::||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      LGKGNNESTKTGMAGSRLLACGVIG 151
|||:::||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 12
S07007
Superoxide dismutase (EC 1.15.1.1) (Cu-Zn) 4, cytosolic [validated] - maize
C:Species: Zea mays (maize)
C>Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 21-Jul-2000
C:Accession: S07007; S20589; S72234
R:Cannon, R.E.; Scandalios, J.G.
Mol. Gen. Genet. 219, 1-8, 1989
A>Title: Two cDNAs encode two nearly identical Cu/Zn superoxide dismutase proteins in
A:Reference number: S07007; MUID:90136495; PMID:2482436
A:Molecule type: mRNA
A:Molecule type: mrna
A:Residues: 1-152 <CAN>
A:Cross-references: EMBL:X17565; NTD:g22483; PIDN:CA857992.1; PID:g618682
A:Experimental source: Inbred line W64A
A:Accession: S20569
A:Molecule type: protein
A:Residues: 2-20 <CAN2>
R:Kernodle, S.P.; Scandalios, J.G.
Genetics 144, 317-328, 1996
A>Title: A comparison of the structure and function of the highly homologous maize an
A:Reference number: S72234; MUID:97032947; PMID:8878655
A:Accession: S72234
A:Molecule type: DNA
A:Residues: 1-124 <KEP>
A:Cross-references: EMBL:U34726; NTD:g1574937; PIDN:AAB49912.1; PID:g1574938
C:Gene: sod4
A>Introns: 26/1; 60/1; 92/1; 102/3
C:Function:
A>Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxyg
C:Superfamily: superoxide dismutase (Cu-Zn)
C:Keywords: copper; cytosol; metalloprotein; oxidoreductase; zinc
F:2-152/Product: superoxide dismutase (Cu-Zn) #status experimental <MAT>
F:45,47,62,119/Binding site: copper (His) #status predicted
F:56-145/Dissulfide bonds: #status predicted
F:62,70,79,82/Binding site: zinc (His, His, His, Asp) #status predicted
F:142/Active site: Arg #status predicted

Query Match          5.9%; Score 164; DB 2; Length 152;
Best Local Similarity 31.7%; Pred.No. 5e-05;
Matches 45; Conservative 21; Mismatches 50; Indels 26; Gaps 4;

Db      14 VKGITPTQGDDPRPVTVGSVSL-----KRLGFRHVALGSTTGCGSKSTGPHTNPAS 67
|||:::||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

```

OY      283 -----EDVHSELGGLGDHDHDCGVNESHRSRWIMIFDDSDVLGRSLAH-QRQNLHK 336
           |||:|||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      68 KENGAPEDEENRRHAGDGLGNVTAGADSVANINIVTDSQILFLEPNSIIGRAVVVNADPDLDLK 127

OY      337 -----SAKIACCVTG 346
           :::::|||:
Db      128 GGHLSKSTGMNAGRVACGIIG 149

RESULT 13
S22508
superoxide dismutase (EC 1.15.1.1) (Cu-Zn) soda - rice
C:Species: Oryza sativa (rice)
C>Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 20-Jun-2000
A:Accession: S22508
R:Sakamoto, A.; Ohnuga, H.; Tanaka, K.
Plant Mol. Biol. 19, 323-327, 1992
A>Title: Nucleotide sequences of two cDNA clones encoding different Cu/Zn-superoxide dis-
F:562.70.79.82/Binding site: zinc (His, His, His, Asp) #status predicted
F:142/Active site: Arg #status predicted
A:Molecule type: mRNA
A:Residues: 1-152 <SAR>
A:Cross-references: EMBL:D00999; NID:G218223; PIDN:BAA00799.1; PID:G218224
A:Experimental source: clone RSODA
C:Genetics:
A:Gene: soda
C:Function:
A>Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxygen
C:Superfamily: superoxide dismutase (Cu-Zn)
C:Keywords: copper; metalloprotein; oxidoreductase; zinc
F:45.47.62.119/Binding site: copper (His) #status predicted
F:56.145/Dissulfide bonds: #status predicted
F:62.70.79.82/Binding site: zinc (His, His, His, Asp) #status predicted
F:142/Active site: Arg #status predicted

Query Match          5.8% Score 163; DB 2; Length 152;
Best local similarity 32.4%; Pred. No. 5,9e-05;
Matches 46; Conservative 19; Mismatches 51; Indels 26; Gaps 4;

OY      225 VHGCTDRKYGYCDPLEVSYHLSEGFNVSDDKRHNDVQITANGDLGGCDMLAKXPR-- 282
           |||:|||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      14 VKGTTHFVQEDRGDTTYTGSGSL-----KRLGFGFIHALCGPTNGCKSTGPNRYRAG 67

OY      283 -----HEDVHSELGGLGDHDHDCGVNESHRSRWIMINIFGDDSVLGSIALH-QRQNLHK 336
           |||:|||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      68 KENGAPEDEETRNHADLCNVATAGEDSVANIHVVDSQILFRGSIIIGRAVVVNADPDLDLK 127

OY      337 -----SAKIACCVTG 346
           :::::|||:
Db      128 GGHLSKSTGMNAGRVACGIIG 149

RESULT 14
S72235
superoxide dismutase (EC 1.15.1.1) (Cu-Zn) 4A, cytosolic [validated] - maize
C:Species: Zea mays (maize)
C>Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 21-Jul-2000
C:Accession: S72235; S78441; S07008; S20570
R:Kernodele, S.P.; Scandalios, J.G.
Genetics 144, 317-328, 1996
A>Title: A comparison of the structure and function of the highly homologous maize anti-
A:Reference number: S72234; WUID:97032947; PMID:8878695
A:Accession: S72235
A:Molecule type: DNA
A:Residues: 1-152 <KER>
A:Cross-references: EMBL:U34727; NID:g1899026
A>Note: the authors translated the codon CAC for residue 61 as Pro
R:Scandalios, J.G.
submitted to the EMBL Data Library, August 1995
A:Reference number: S78441
A:Accession: S78441
A:Molecule type: DNA
A:Residues: 1-61, D_63-152 <SCA>
A:Residues: 1-61, D_63-152 <SCA>
```

A:Cross-references: EMBL:U34727; NID:g1899026; PIDN:AAB49913.1; PID:g1885354
R:Gannon, R.E.; Scandilios, J.G.
Mol. Gen. Genet. 219, 1-8, 1989

A>Title: Two cDNAs encode two nearly identical Cu/Zn superoxide dismutase proteins in
A:Reference number: S07007; MUID:90136495; PMID:2482436

A:Accession: S07008

A:Molecule type: mRNA

A:Residues: 1-61, 'O', '63-152 <CAN>

A:Cross-references: EMBL:X17564; NID:g22484; PIDN:CAB57993.1; PID:g6018746

A:Experimental source: Inbred line W64A

A>Note: The authors translated the codon CAC for residue 61 as Pro and GAC for residu

A:Molecule type: protein

A:Residues: 2-20 <CANZ>

C:Genetics:

A:Gene: sod4A

A:Introns: 26/1; 60/1; 92/1; 102/3; 128/1; 146/1

C:Function:

A>Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxyg

C:Superfamily: superoxide dismutase (Cu-Zn)

C:Keywords: copper; cytosol; metalloprotein; oxidoreductase; zinc

F:2-152/Product: superoxide dismutase (Cu-Zn) #status experimental <MAT>

F:45,47,62,119/Binding site: copper (His) #status predicted

F:56-145/Disulfide bonds: #status predicted

F:62,70,79,82/Binding site: zinc (His, His, Asp) #status predicted

F:142/Active site: Arg #status predicted

Query Match 5.8%; Score 163; DB 2; Length 152;
Best Local Similarity 31.7%; Pred. No. 5,9e-05;
Matches 45; Conservative 21; Mismatches 50; Indels 26; Gaps 4;

Oy 225 VHGTFDKQGVYDDELVSYLEGFVNSDDHDLADVOIYANGDLTSGCDNIQAYDP-- 282
| | | | | : : : : :
Db 14 VKGIFFTCGGDPRVTAVTGSVGL-----KFGIGHFVHALDDITNGCMTSGHHYNPAS 67
| | | | | : : : : :
Oy 283 -----HDVYSSEIGDGDIIHDDHGVNESHRYSWINIFPDSDSVLGSRSLAH-QRDHLHK 336
| | | | | : : : : :
Db 68 KKHGAPEDENRHAGDIGNTACADGVANINVTDSQILPTGPSNIIIRAVVVHADPDDL GK 127
| | | | | : : : : :
Oy 337 -----SAKIACCVI G 346
| | | | | : : : : :
Db 128 GGHELSKSTGNAGGRVACGIIG 149

RESULT 15

J00915

Superoxide dismutase (EC 1.15.1.1) (Cu-Zn) - mouse

N.Alternate names: proteasome chain CB

C.Species: Mus musculus (house mouse)

C.Date: 30-Sep-1991 #sequence revision 30-Sep-1991 #text_change 20-Apr-2000

C.Accession: J00915; S00910; A37101

R.Benedetto, M.L.; Anzal, Y.; Gordon, J.W.

Gene 99, 191-195, 1991

A>Title: Isolation and analysis of the mouse genomic sequence encoding Cu2+-Zn2+ supe

A:Reference number: J00915; MUID:91216458; PMID:2022332

A:Accession: J00915

A:Molecule type: DNA

A:Residues: 1-154 <BEN>

A:Cross-references: GB:M60794

A.Note: The authors translated the codon CAT for residue 102 as Asp

A>Note: The sequence shown follows the authors' translation

R:Bewley, G.C.

Nucleic Acids Res. 16, 2728, 1988

A>Title: cDNA and deduced amino acid sequence of murine Cu-Zn superoxide dismutase.

A:Reference number: S00910; MUID:88203220; PMID:3362653

A:Accession: S00910

A:Molecule type: mRNA

A:Residues: 1-154 <BEN>

A:Cross-references: EMBL:X06683; NID:g54127; PIDN:CAA29880.1; PID:g54128

R:Plutheo, F.G.; Shreeve, M., Ecklinazi, D.; van der Gaag, H.; Huang, K.S.; Hulmes, J

J. Cell Biol. 111, 1217-1223, 1990

A>Title: Purification of an inhibitor of erythroid progenitor cell cycling and antago

A:Reference number: A37101; MUID:90361747; PMID:2391363
 A:Accession: A37101
 A:Molecule type: protein
 A:Residues: 4-6, 'X', 8-23 <PLU>
 C:Genetics:
 A:Gene: SOD-1
 A:Introns: 24/3; 57/1; 80/2; 119/3
 C:Function:
 A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxygen
 C:Superfamily: superoxide dismutase (Cu,Zn)
 C:Keywords: copper metalloprotein; oxidoreductase; zinc
 F:47/49/64,121/Binding site: copper (His) #status predicted
 F:58-147/Disulfide bonds: #status predicted
 F:64,72,81,84/Binding site: zinc (His, His, His, Asp) #status predicted
 F:144/Active site: Arg #status predicted

Query Match 5.8%; Score 162.5; DB 2; Length 154;
 Best Local Similarity 30.1%; Pred. No. 6.5e-05;
 Matches 44; Conservative 25; Mismatches 50; Indels 27; Gaps 6;

QY 225 VHGTFKQVGYD-LEVSYHLEGFNVSDDKDHLHDVQIYANGDLTSCGDNLGAKYDPH 283
 Db 15 VOGTIFHEOKASGEPVLSGQITGL-TEGQHGPHQY-----GDNTQCTSGAPHPNPH 68
 QY 284 EDYHS-----ELGDLGDIHDDHGVNESHRYSWINIFGDDSVLGRSIAIHQR-DHLH 335
 Db 69 SKKHGPADEERHVGDLGNTAGKGVANVSIEDRVISTSGEHSIIIGRTWVYHEKQDDLG 128
 QY 336 K-----SAKINCCVYGRGO 349
 Db 129 KCGNEESTKTNAGSRIACGVIGITNQ 154

Search completed: November 29, 2002, 12:36:07
 Job time : 49 secs

THIS PAGE BLANK (USPTO)


```

Db 67 EENNERNNEENNERNNEENNERNNNRRNN-----NNNNRRNNNNLGNH--NH 120
Oy 71 ELILUGFNTESEDDNNGLNLMLGMSASGDSIGELUNAPRKNDROLDLDVDDOR 130
Db 121 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 152
Oy 131 GVVNENYUMLDIDSTARTALGSKMTILOSSTADPRARASICVIGSKARPE 190
Db 153 ----NANNNNWF-----NNNLGNNNNNARNNNNNNNNNNNNNNNNNNNNNNN 195
Oy 191 AALNLELEDKTEHACDVRSNTORKALNNYKSTIDTKOYGVDELVSYHLSGFV 250
Db 196 ----NNNNNNNARNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 245
Oy 251 SDDKHOLDVQIANGDLTSGCDNLGAKYURREDYHSELDDLDDDDHGVVNESHR 310
Db 246 --NNNNNNNDAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 282
Oy 311 SWINFGDSDVLGRSTALNQRDLKRSKACVIGRQSHREIVNRAKCVAREPTESTG 370
Db 283 D-----ANNNNNNNDA-----NNNNNNNNNNNNNNNNNNNNNNNNNNNN 295
Oy 371 LNNHVSIGITFEQTPGSGTNTADLKGFNVEDLSNNHNGVQLHEWGDMSHCHSLGRMY 430
Db 296 -NNN-----NNNNDAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 337
Oy 431 HGNDAN 437
Db 338 NNNNDAN 344

RESULT 2
SRCH_RABIT STANDARD: PRT: 852 AA.
AC PIC230;
Db 01-APR-1990 (Rel. 14, Created)
Db 01-APR-1990 (Rel. 14, Last sequence update)
Db 16-OCT-2001 (Rel. 40, Last annotation update)
Db Sarcoplasmic reticulum histidine-rich calcium-binding protein
DE precursor.
GN HRC OR HCP.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC TISSUE=Skeletal muscle;
RX MEDLINE=90036884; PubMed=2808365;
RA Hofmann S.L., Goldstein J.L., Orth K., Moosaw C.R., Slaughter C.A.,
RA Brown M.S.;
RT "Molecular cloning of a histidine-rich Ca2+-binding protein of
RT sarcoplasmic reticulum that contains highly conserved repeated
RT elements.";
RT J Biol Chem. 264:18083-18090(1989).
CC -1- FUNCTION: MAY PLAY A ROLE IN THE REGULATION OF CALCIUM
CC SQUESTRATION OR RELEASE IN THE SR OF SKELETAL AND CARDIAC MUSCLE.
CC -1- SUBCELLULAR LOCATION: SARCOPLASMIC RETICULUM LUMEN.
CC -1- SIMILARITY: STRONG, TO HUMAN HRC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J05080; AAA31279.1; -
DR PIR: A34373; A34373
DR InterPro: IPR002134; HCP
DR PROSITE: PS00328; HCP; 10
KM Calcium-binding; Signal; Repeat.

```

```

FT SIGNAL 1 27
FT CHAIN 28 852
FT MOD_RES 28 28
FT DOMAIN 59 100
FT REPEAT 59 79
FT REPEAT 80 100
FT DOMAIN 199 470
FT REPEAT 199 224
FT REPEAT 225 253
FT REPEAT 254 282
FT REPEAT 283 310
FT REPEAT 311 339
FT REPEAT 340 367
FT REPEAT 368 395
FT REPEAT 396 423
FT REPEAT 424 451
FT REPEAT 452 471
FT DOMAIN 471 585
FT DOMAIN 721 826
FT SEQUENCE 852 AA; 96117 MW; A43C0CB3494B930 CRC64;

Query Match 6.4%; Score 178; DB 1; Length 852;
Best local similarity 22.2%; Pred. No. 8.5e-06;
Matches 126; Conservative 54; Mismatches 176; Indels 212; Gaps 38;

Oy 1 DGEQCNQGNKNDHNDH-----NDHNDHNDHDDDTMTYACSEMRPNMASSLNNHNG 56
Db 237 DDDDEGSDTESDNNHNNRRNGHEEDEDDE-----GDSTESDRQA-----NRN--- 285
Oy 57 SIELSKGH-----GAVYLELILV-----GFNTSEDDND-----NNHGL 90
Db 286 -----RGHREEREDDDDEGSDTESDRQAHRNGHREEDDDDDDEGSDTESDRQA 339
Oy 91 HLNH-----LDDMSAGDSIGELYNH-----PEKNAPRGELGVDDRGVYEVY 137
Db 340 HRNGHREEDDDDDDEGSDTESDRQAHRNGHREEDDDDDDEGSDTESDR-----NQHN 396
Oy 138 HYAN-----LDIDGAPYTEALIGHSMIL-----QGSHTDPT-PAST 176
Db 397 RNRHREEDDDDDDEGSDTES-----DHNQANHRNGHREEDDDDDDEGSDTESDRQAHRN 454
Oy 177 ACSYIGHKARPE--TAAALNHELE-----EDKTE--HYAN----- 208
Db 455 R-----GHREEREDDDDEGSDTESDRQAHRNGHREEDDDDDDDGSDTESGNAHRQNG 510
Oy 209 ---CDVRSNTNORKALNNHY--HGTIDFKOVGYDLEVSYNLGEFNVSDHKNLNDVOI 263
Db 511 KEELVTSDEH-----NNHVPRHG-----HOGHGDKKEE--EGVST-----DHMQVPR 553
Oy 264 YAN-----GDLTSGCDNLGAKYURREDYHSELGLDGL-DIHDDHGVV-----NESHRYSWINT 315
Db 554 NANNRGEEGEGEELTYKAGNNHVASRPRGRSREGHAEHOTOEVGNHOH----- 607
Oy 316 FGDDSVLGRSTALNODNHLKSKACVIGRQSHREIVNRAKCVAREPTESTGLNNHY 375
Db 608 -----MGDDTSAENGRASSP-----RQGNH-----REDY-VNNH-- 638
Oy 376 SGSTFE---QTG-----GSTNTADLKGFNVEDLSNNHNGVQLHEWGDMS 420
Db 639 RGSLEKEVGRSPRGVAKGDSVYKSGSEEEEDKG-----TNN-HSLDEDEBEG 650
Oy 421 HGCHSLGRMYHGNDANHPKRPGLDGV 448
Db 691 HG-RSLSO-----EDDEEDRGEESAKV 712

RESULT 3
SODC_HALRO STANDARD: PRT: 151 AA.
ID SODC_HALRO
AC P81926;
DT 30-MAY-2000 (Rel. 39, Created)

```

```

DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Superoxide dismutase [Cu-Zn] (EC 1.15.1.1).
OS Halocynthia roretzi (Sea squirt).
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
OC Stolidobranchia; Pyrosidae; Halocynthia.
OX NCBI_TaxID=7729;
RN [1]
RP SEQUENCE, AND CHARACTERIZATION.
RC TISSUE-Hemocyte, and Plasma;
RX MEDLINE=99302489; PubMed=10374259;
RA Abe Y., Ishikawa G., Satoh H., Azumi K., Yokosawa H.;
RT Primary structure and function of superoxide dismutase from the
RT ascidian Halocynthia roretzi.
RL Comp. Biochem. Physiol. 122B:321-326(1999).
CC -1- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems. The plasma
CC superoxide dismutase has phagocytosis-stimulating activity and may
CC play an important role in the biological defenses of the organism.
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -1- COFACTOR: Copper and zinc (By similarity).
CC -1- ENZYME REGULATION: INHIBITED BY KCN AND DIETHYLDITHIOCARBAMATE.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- PTM: ACETYLATED N-TERMINAL SERINE.
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
DR HSSP: P15107: 1XSO.
DR InterPro: IPR001424: SOD_CU_ZN.
DR Pfam: PF00080: sodcu.1.
DR PRINTS: PR00068: CUZNDISMUTASE.
DR ProDom: PD000469: SOD_CU_ZN_1.1.
DR PROSITE: PS00087: SOD_CU_ZN_1.1.
DR PROSITE: PS00332: SOD_CU_ZN_2; FALSE_NEG.
KM Oxidoreductase; Copper; zinc.
FT METAL 43 43 COPPER (BY SIMILARITY).
FT METAL 45 45 COPPER AND ZINC (BY SIMILARITY).
FT METAL 60 60 COPPER AND ZINC (BY SIMILARITY).
FT METAL 68 68 ZINC (BY SIMILARITY).
FT METAL 77 77 ZINC (BY SIMILARITY).
FT METAL 80 80 ZINC (BY SIMILARITY).
FT METAL 117 117 COPPER (BY SIMILARITY).
FT METAL 143 143 COPPER (BY SIMILARITY).
SQ DISULFID 54 143 BY SIMILARITY.
SQ SEQUENCE 151 AA; 15489 MW; 162F81A82275AF0 CRC64;

Query Match 6.3%; Score 175; DB 1; Length 151;
Best Local Similarity 29.6%; Pred. No. 1.6e-06;
Matches 42; Conservative 26; Mismatches 48; Indels 26; Gaps 5;

QY 225 VHGCTIDFKVGYGDLVSYHLEGFVNSDDKRDHLADVOIYANGDLTSGCDNLGAKYDPHE 284
| | | | | : : : | | | : : : | | | : : : | | | : : : |
DB 12 VKGTLNFKODALIGCTVGEVSGL-IPGKHGFHIEY-----GDLNFGCTSSGCHFNPK 65
| | | | | : : : | | | : : : | | | : : : | | | : : : |
QY 285 DYHS-----ELGDLGDIHDDHGVNESHRYSMINIFGDDSVLGRSIAHQ-RDHL-- 334
| | | | | : : : | | | : : : | | | : : : | | | : : : |
DB 66 QIHGAREDDIRHVGDGLNTADSSGAVATNITDMISLTGESHISIGRAVVVHAGEDDLGK 125
| | | | | : : : | | | : : : | | | : : : | | | : : : |
QY 335 -----HRSKTIACVIG 346
| | | | | : : : | | | : : : | | | : : : | | | : : : |
DB 126 GGHEDSKTGHAGRLSGVIG 147
| | | | | : : : | | | : : : | | | : : : | | | : : : |

RESULT 4
SODC_PEA
ID SODC_PEA STANDARD: PRT: 151 AA.
AC 002610:
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1999 (Rel. 26, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Superoxide dismutase [Cu-Zn] (EC 1.15.1.1).
GN SODC.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

```

```

OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RA White D.A., Zilinskas B.A.;
RT "Nucleotide sequence of a complementary DNA encoding pea cytosolic
RT copper/zinc superoxide dismutase."
RL Plant Physiol. 96:1391-1392(1991).
CC -1- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -1- COFACTOR: Copper and zinc.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M63003: AAA3659.1; -.
DR HSSP: P07505: 1SRD.
DR InterPro: IPR001424: SOD_CU_ZN.
DR Pfam: PF00080: sodcu.1.
DR PRINTS: PR00068: CUZNDISMUTASE.
DR ProDom: PD000469: SOD_CU_ZN_1.1.
DR PROSITE: PS00087: SOD_CU_ZN_1.1.
DR PROSITE: PS00332: SOD_CU_ZN_2; 1.
KM Oxidoreductase; Copper; zinc.
FT INITMET 0 0 BY SIMILARITY.
FT METAL 44 44 COPPER (BY SIMILARITY).
FT METAL 46 46 COPPER AND ZINC (BY SIMILARITY).
FT METAL 61 61 COPPER AND ZINC (BY SIMILARITY).
FT METAL 69 69 ZINC (BY SIMILARITY).
FT METAL 78 78 ZINC (BY SIMILARITY).
FT METAL 81 81 ZINC (BY SIMILARITY).
FT METAL 118 118 COPPER (BY SIMILARITY).
FT METAL 144 144 COPPER (BY SIMILARITY).
SQ DISULFID 55 144 BY SIMILARITY.
SQ SEQUENCE 151 AA; 15192 MW; 9BCFA53DD2CB1719 CRC64;

Query Match 6.2%; Score 173.5; DB 1; Length 151;
Best Local Similarity 31.4%; Pred. No. 2.1e-06;
Matches 49; Conservative 26; Mismatches 48; Indels 33; Gaps 5;

QY 211 VRSNTQPKALHNHAGTIDFKVGYGDLVSYHLEGFVNSDDKRDHLADVOIYANGDLT 270
| | | | | : : : | | | : : : | | | : : : | | | : : : |
DB 6 VLSNSNE-----VSGTINFSQEGNGPTVYGLAGL-----KFLGPHIHALGDTT 52
| | | | | : : : | | | : : : | | | : : : | | | : : : |
QY 271 SGCDNLGAKYDPH-----EDYHSELGDLGDIHDDHGVNESHRYSMINIFGDDSVLG 323
| | | | | : : : | | | : : : | | | : : : | | | : : : |
DB 53 NGCISTGPHFNNGKHEGAPDETRHAGDLGNINVGDDGVSTPTIDNHIPLTGINSIIG 112
| | | | | : : : | | | : : : | | | : : : | | | : : : |
QY 324 RSTAIR-ORDHLK-----SAKTIACVIG 346
| | | | | : : : | | | : : : | | | : : : | | | : : : |
DB 113 RAVVYHADPDDLKGKGHLSKTGNGAGRAVACIG 148
| | | | | : : : | | | : : : | | | : : : | | | : : : |

RESULT 5
SODE_RAT
ID SODE_RAT STANDARD: PRT: 244 AA.
AC 008420: Q64667.
DT 01-OCT-1994 (Rel. 30, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Extracellular superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1)
DE (EC-SOD) (Superoxide dismutase B).
GN SOD3 OR SOD-3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryonic;
 RX MEDLINE=93319511; PubMed=6328962;
 RA Perry A.C.F., Jones R., Hall L.;
 RT "Isolation and characterization of a rat cDNA clone encoding a
 RT secreted superoxide dismutase reveals the epididymis to be a major
 RT site of its expression."
 RL Biochem. J. 293:21-25(1993).
 RN [2]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RX MEDLINE=94043314; PubMed=8227019;
 RA Williams J., Zwijsen A., Slegers H., Nicolai S., Bettadapura J.,
 RA Raymakers J., Scarce T.;
 RT "Purification and sequence of rat extracellular superoxide dismutase
 RT B secreted by C6 glioma."
 RL J. Biol. Chem. 268:24614-24621(1993).
 RN [3]
 RP SEQUENCE FROM N.A., AND MUTAGENESIS.
 RC STRAIN=Sprague-Dawley; TISSUE=Uterus;
 RX MEDLINE=96224263; PubMed=8643556;
 RA Carlson L.M., Marklund S.M., Edlund T.;
 RT "The rat extracellular superoxide dismutase dimer is converted to a
 RT tetramer by the exchange of a single amino acid."
 RL Proc. Natl. Acad. Sci. U.S.A. 93:5219-5222(1996).
 CC -1- FUNCTION: Destroys radicals which are normally produced within the
 CC cells and which are toxic to biological systems.
 CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -1- COPFACTOR: Copper and zinc.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X68041; CAA48177.1; -;
 DR EMBL: Z24721; CAA80849.1; -;
 DR EMBL: X94371; CAA64149.1; -;
 DR PIR: S34055; S34055.
 DR HSSP: P00441; LSOS.
 DR InterPro: IPR001424; SOD_CU_ZN.
 DR Pfam: PF000080; sdcu.1.
 DR PRINTS: PD00068; CUZNDISMUTASE.
 DR PRODOM: PD000469; SOD_CU_ZN.1.
 DR PROSITE: PS00087; SOD_CU_ZN.1;
 DR PROSITE: PS00332; SOD_CU_ZN.2; 1.
 KM Oxidoreductase; Copper; Zinc; Glycoprotein; Signal.
 FT SIGNAL 1 15
 FT CHAIN 16 244
 FT FT
 FT METAL 121 121
 FT METAL 123 123
 FT METAL 138 138
 FT METAL 146 146
 FT METAL 149 149
 FT METAL 152 152
 FT METAL 188 188
 FT METAL 214 214
 FT DISULFID 132 214
 FT CARBOHYD 114 114
 FT MURAGEN 48 48
 FT CONFLICT 235 235
 FT CONFLICT 237 237
 FT SEQUENCE 244 AA; 26620 MW; B66726501CE58614 CAC64;
 Query Match 6.28; Score 173.5; DB 1; Length 244;

Best Local Similarity 29.7%; Pred. No. 3.8e-06;
 Matches 54; Conservative 28; Mismatches 61; Indels 39; Gaps 10;
 QY 194 LHHLEEDKTEHNAKCDYRSNTHPKAL-----HHNVKGTDFKQVGGD-LEVSXHEGF 248
 DB 55 LGKQREADARKEHNAKRV-----QPSAMLPDQPOITGLVLEFRLGSSRLKASFNLEGF 109
 QY 249 NVSDHDKDHLHDVQIYANGDLSGCDNLGAKYDPHEHYHSELGLDGIHDDHG--VYNE 306
 DB 110 PAEONTSNHA--IHVEFEGLSGGCESTSPHNP-----LGVPHDPHDEFEFNVRD 160
 QY 307 ---SHRYSW-INIFGDDSVLGRSIAIHQ-RDHLKSA-----KIACVYGRG 348
 DB 161 GLMKHRRGLATSLAGPHSITLGRAVYVHAGEDDLKGGNGASVONGNAGRRLACCVYGT 220
 QY 349 QS 350
 DB 221 NS 222
 RESULT 6
 SODE_MOUSE
 ID SODE_MOUSE STANDARD; PRI; 251 AA.
 AC 009164;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Extracellular superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1)
 DE (EC-SOD).
 GN SOD3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-36.
 RC STRAIN=C57BL/6 J CBA; TISSUE=Lung;
 RX MEDLINE=98026284; PubMed=9376114;
 RA Folz R.J., Guan J., Seldin M.F., Oury T.D., Englund J.J., Crapo J.D.;
 RT "Mouse extracellular superoxide dismutase: primary structure, tissue-
 RT specific gene expression, chromosomal localization, and lung in situ
 RT hybridization."
 RL Am. J. Respir. Cell Mol. Biol. 17:393-403(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Kidney;
 RX MEDLINE=97306432; PubMed=9163733;
 RA Oh Y.-S., Wada K., Yamashita T., Kikuchi T., Folz R.J., Tanaka K.;
 RT "Sequence analysis, tissue expression and chromosomal localization of
 RT a mouse secreted superoxide dismutase gene."
 RL Mol. Cells 7:204-207(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Liver;
 RA Siegfried M.R., Schultz D., Harrison D.G., Fukui T.;
 RT "Murine extracellular superoxide dismutase genomic sequence."
 RT Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Destroys radicals which are normally produced within the
 CC cells and which are toxic to biological systems.
 CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -1- COPFACTOR: Copper and zinc (By similarity).
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----

```

DR EMBL: U38261; AAB51106.1; -
DR EMBL: D50856; BAA23493.1; -
DR EMBL: AF223251; AAF27932.1; -
DR HSSP: P00445; 1JCV.
DR MGD: MGI:103181; SOD3.
DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; sodcu.1.
DR PRINTS: PR00068; CUZNDISMUTASE.
DR ProDom: PD000469; SOD_CU_ZN.1.
DR PROSITE: PS00087; SOD_CU_ZN.1.
DR PROSITE: PS00332; SOD_CU_ZN.2.1.
KW Oxidoreductase; Copper; zinc; Glycoprotein; Signal.
FT SIGNAL 1 15
FT PROPEP 16 24
FT CHAIN 25 251
FT METAL 128 128 EXTRACELLULAR SUPEROXIDE DISMUTASE [CU-
FT METAL 130 130 ZN].
FT METAL 145 145 COPPER (BY SIMILARITY).
FT METAL 153 153 COPPER AND ZINC (BY SIMILARITY).
FT METAL 156 156 ZINC (BY SIMILARITY).
FT METAL 159 159 ZINC (BY SIMILARITY).
FT METAL 195 195 ZINC (BY SIMILARITY).
FT DISULFID 121 121 BY SIMILARITY.
FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 251 AA; 27392 MW; 9EAF850E458966C4 CRC64;

Query Match 6.1%; Score 171.5; DB 1; Length 251;
Best Local Similarity 30.6%; Pred. No. 5.6e-06;
Matches 55; Conservative 23; Mismatches 57; Indels 45; Gaps 11;

OY 199 EEKRTGTHVHCVNRNTQPKAL---HHVHGTFDKQVGYD-LEVSYLEGCVNSD 253
DB 67 EYDAAMHAIICRV-----QPSATLEPPDQPIGTGLVLEFRLQGPSRLAEAFSLGPFABEN 121
OY 254 HKD---HLHDVQIYANGTISGCCNLGKXPHEDVSELDGDIHDDHG--VYNES- 307
DB 122 ASRAIAHVEF-----GDSGOCDSGTGHVPMVPPV-----HPSDFGFFVNRNQ 169
OY 308 ---HRS-WINRPGDVSIGSAIHQ-RDLHNSA-----KIACCYIRGQS 350
DB 170 LMRHRYGLTASLAGFHALGHSVYHAGEDLDGKGNQASLQNGNRRLACCVGTSSS 229

RESULT 7
SODE_MAIZE
ID SODE_MAIZE STANDARD; PRT; 150 AA.
AC P11428;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Superoxide dismutase [Cu-Zn] 2 (EC 1.15.1.1).
GN SODCC.1 OR SOD2.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_Taxid=4577;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=87092359; Pubmed=3467349;
RA Cannon R.E., White J.A., Scandalios J.G.;
RT "Cloning of cDNA for maize superoxide dismutase 2 (SOD2).";
RL Proc. Natl. Acad. Sci. U.S.A. 84:179-183(1987).
RN 12
RP SEQUENCE FROM N.A.
RX MEDLINE=87249683; Pubmed=3597043;
RA Cannon R.E., Scandalios J.G.;
RT "The superoxide dismutase-2 gene of maize.";
RL Isozymes Curr. Top. Biol. Med. Res. 14:73-81(1987).
CC 1- FUNCTION: Destroy radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC 1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).

```

```

CC 1- CORFACTOR: Copper and zinc (by similarity).
CC 1- SUBUNIT: HOMODIMER.
CC 1- SUBCELLULAR LOCATION: Cytoplasmic.
CC 1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
CC
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: M15175; AAA3511.1; -
DR EMBL: M54936; AAA3510.1; -
DR PIR: A29077; A29077.
DR HSSP: P07505; 1SRD.
DR MaizeDB: 47586; -.
DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; sodcu.1.
DR PRINTS: PR00068; CUZNDISMUTASE.
DR ProDom: PD000469; SOD_CU_ZN.1.
DR PROSITE: PS00087; SOD_CU_ZN.1.
DR PROSITE: PS00332; SOD_CU_ZN.2.1.
KW Oxidoreductase; Copper; zinc; Multigene family.
FT INIT MET 0 0
FT METAL 43 43 COPPER (BY SIMILARITY).
FT METAL 45 45 COPPER (BY SIMILARITY).
FT METAL 60 60 COPPER AND ZINC (BY SIMILARITY).
FT METAL 68 68 ZINC (BY SIMILARITY).
FT METAL 77 77 ZINC (BY SIMILARITY).
FT METAL 80 80 ZINC (BY SIMILARITY).
FT METAL 117 117 COPPER (BY SIMILARITY).
FT DISULFID 54 143 BY SIMILARITY.
SQ SEQUENCE 150 AA; 14973 MW; C63A117072C37873 CRC64;

Query Match 6.0%; Score 168; DB 1; Length 150;
Best Local Similarity 31.9%; Pred. No. 5.3e-06;
Matches 46; Conservative 22; Mismatches 46; Indels 30; Gaps 5;

OY 225 VHGTFDKQVGYDLEVSYLEGCVNSDHRKDLHDVQIYANGTISGCCNLGKXPHEDVSELDGDIHDDHG--VYNES- 282
DB 12 VAGTLEFSEQEGDGPPTVIGSISGL-----KPGDHGHVHALDSDTNGSMSTGPHFNPVG 65
OY 283 -----HEDVSELDGDIHDDHGCVNESHRYSWINIFGDSVIGRSIAIH-QRDHL 334
DB 66 KEHGAEDEDRHA--GDLGNTYAGEDGVVYNNITDSDIPLAGHSITIGRAVYVHADPDL 123
OY 335 HK-----SAKIACCYIG 346
DB 124 GKQGHLSKSTGNACGRVACGII 147

RESULT 8
SODE_RABBIT
ID SODE_RABBIT STANDARD; PRT; 244 AA.
AC P41975;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Extracellular superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1)
DE (EC-SOD).
GN SOD3.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_Taxid=9986;
RN 11
RP SEQUENCE FROM N.A.
RX STRAIN=New Zealand white;
RA Laukkanen M.O., Aittomaki S.J., Hiltunen T.P., Yla-Herttuala S.;
RL submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN 12

```

RP SEQUENCE FROM N.A.
RC STRAIN-New Zealand white; TISSUE-Heart;
RA Laakkonen M.O., Hiltunen M.O., Altomaki S., Janne J.,
RA Yla-Herttuala S.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Laakkonen M.O., Altomaki S., Mannermaa S., Hiltunen M.O.,
RA Yla-Herttuala S.;
RT Cloning and characterization of rabbit extracellular superoxide
RT dismutase.*;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 43-155 FROM N.A.
RC STRAIN-New Zealand white; TISSUE-Aorta;
RA Hiltunen T.P., Nikkari T., Yla-Herttuala S.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC - FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC - CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) - O(2) + H(2)O(2).
CC - COPROCTOR: Copper and zinc (By similarity).
CC - SUBUNIT: HOMODIMER (BY SIMILARITY).
CC - SUBCELLULAR LOCATION: Extracellular.
CC - SIMILARITY: BELONGS TO THE CO-ZN SUPEROXIDE DISMUTASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z67878; CAA91785.1; -
DR EMBL: Y13339; CAA73783.1; -
DR EMBL: AJ007044; CAA07431.1; -
DR EMBL: X78139; CAA55018.1; -
DR HSSP: P00442; ICBJ.
DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; SODCU.1.
DR PRINTS: PR00068; CUZNDISMUTASE.
DR PRODOM: PD000469; SOD_CU_ZN.1.
DR PROSITE: PS00087; SOD_CU_ZN.1; 1.
DR PROSITE: PS00332; SOD_CU_ZN.2; 1.
KW Oxidoreductase; Copper; zinc; Glycoprotein; Signal.
FT SIGNAL 1 18
FT CHAIN 19 244
FT METAL 118 118
FT METAL 120 120
FT METAL 135 135
FT METAL 143 143
FT METAL 146 146
FT METAL 149 149
FT METAL 185 185
FT DISULFID 129 211
FT CARBOHYD 111 111
FT CORRECT 53 53
SQ SEQUENCE 244 AA; 25668 MW; 7C9B1C5994FE25 CRC64;
Query Match 5 0%; Score 167.5; DB 1; Length 244;
Best Local Similarity 29.2%; Pred. NO. 1.1e-05;
Matches 50; Conservative 19; Mismatches 67; Indels 35; Gaps 7;
DB 349 QSPHETVHAKCVYRPNTSTGLHHVSGSTFEDT-PGSGTHMTADLKGFNNSSEDLSHH 407
DB 56 QGEFAGALHAVCRVPSATLDAQPRVSLVFRQLGAGOLEAFLEDFGFPVEANLS-- 113
DB 408 RHGVQLHEMGDSHGHSLGRMYGHGDAHDKRRDGLADV-----ID 450
DB 114 SRAIHVHQFDLQGGCDSTGANHNPAAVQH-PQHGDEGNEAVNRGRIMKRSGLAASLA 172
DB 451 DSHGIVSTFTFDHLNVEDLNARSLVIMQGHVESE-----RVACCVIG 495

DB 173 GPHSIV-GRAYVYVHAGEDDLG-----RCGNMAYENGNACNAPRLACCVYG 215
RESULT 9
ID SODC_YEAST STANDARD; PRT; 153 AA.
AC P00445;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Superoxide dismutase [Cu-Zn] (EC 1.15.1.1).
GN SOD1 OR YJR104C OR J1968.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-88263032; PubMed-3290902;
RA Berningham-McDonogh O., Gralla E., Valentine J.;
RT "The copper, zinc-superoxide dismutase gene of Saccharomyces
RT cerevisiae: cloning, sequencing, and biological activity.*";
RL Proc. Natl. Acad. Sci. U.S.A. 85:4789-4793(1988).
RN [2]
RP SEQUENCE FROM N.A.
RA Remezan Red M., Kirchath L., Hollenberg C.P.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE.
RA Johansen J.T., Overballe-Petersen C., Martin B., Hasemann V.,
RA Svendsen I.;
RT "The complete amino acid sequence of copper, zinc superoxide dismutase
RT from Saccharomyces cerevisiae.*";
RL Carlsberg Res. Commun. 44:201-217(1979).
RN [4]
RP MEDLINE-80227835; PubMed-6993479;
RA Stehman H.M.;
RT "The amino acid sequence of copper-zinc superoxide dismutase from
RT bakers' yeast.*";
RL J. Biol. Chem. 255:6758-6765(1980).
RN [5]
RP SEQUENCE OF 1-10.
RC STRAIN-X2180-1A;
RA Fruhiger S., Hughes G.J., Sanchez J.-C., Hochstrasser D.F.;
RL Submitted (FEB-1996) to the SWISS-PROT data bank.
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE-92126276; PubMed-1772629;
RA Djionovic K., Gatti G., Coda A., Antolini L., Pelosi G., Desideri A.,
RA Falconi M., Marmocchi F., Rotilio G., Bolognesi M.;
RT "Structure solution and molecular dynamics refinement of the yeast
RT Cu,Zn enzyme superoxide dismutase.*";
RL Acta Crystallogr. B 47:918-927(1991).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE-9229261; PubMed-1602482;
RA Djionovic K., Gatti G., Coda A., Antolini L., Pelosi G., Desideri A.,
RA Falconi M., Marmocchi F., Rotilio G., Bolognesi M.;
RT "Crystal structure of yeast Cu,Zn superoxide dismutase.
RT Crystallographic refinement at 2.5-A resolution.*";
RL J. Mol. Biol. 225:791-809(1992).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
RX MEDLINE-99152006; PubMed-10026301;
RA Hart P.D., Baldwin M.B., Ogihara N.L., Nerstisian A.M., Weiss M.S.,
RA Valentine J.S., Eisenberg D.;
RT "A structure-based mechanism for copper-zinc superoxide dismutase.*";
RL Biochemistry 38:2167-2178(1999).
CC - FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC - CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) - O(2) + H(2)O(2).

```

CC -1- COFACTOR: Copper and zinc.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, J03279; AAA34543.1; -.
DR EMBL, Z49604; CA89634.1; -.
DR PIR, A36171; DSHCZ.
DR PDB: 1SDY; 31-JAN-94.
DR PDB: 1YSO; 10-JUN-96.
DR PDB: 1JCV; 08-MAR-96.
DR PDB: 2JCM; 08-JUN-99.
DR PDB: 1B4L; 23-DEC-99.
DR PDB: 1B4T; 23-DEC-99.
DR SWISS-2DPAGE: P00445; YEAST.
DR SGI: S003865; SOD1.
DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; sodcu.1.
DR PRINTS: PR00068; CUZNDISMASE.
DR PRODOM: PD000469; SOD_CU_ZN.1.
DR PROSITE: PS00087; SOD_CU_ZN.1; 1.
DR PROSITE: PS00332; SOD_CU_ZN.2; 1.
KW Oxidoreductase; Copper; Zinc; 3D-structure.
FT INIT MET 0
FT METAL 46 COPPER.
FT METAL 48 COPPER.
FT METAL 63 COPPER AND ZINC.
FT METAL 71 ZINC.
FT METAL 71 ZINC.
FT METAL 80 ZINC.
FT METAL 83 ZINC.
FT METAL 120 ZINC.
FT DISULFID 57 146 COPPER.
FT CONFLICT 55 55 N -> D (IN REF. 4).
FT CONFLICT 92 92 N -> D (IN REF. 4).
FT STRAND 2 8
FT STRAND 14 20
FT STRAND 28 35
FT STRAND 41 48
FT STRAND 54 57
FT STRAND 58 60
FT STRAND 63 63
FT STRAND 66 67
FT STRAND 74 75
FT STRAND 81 82
FT STRAND 83 89
FT STRAND 91 92
FT STRAND 95 101
FT STRAND 105 105
FT STRAND 109 110
FT STRAND 111 111
FT STRAND 113 114
FT STRAND 116 119
FT STRAND 126 127
FT STRAND 133 133
FT STRAND 134 137
FT STRAND 138 139
FT STRAND 145 148
FT STRAND 150 151
SQ SEQUENCE 153 AA; 15723 MW; 4B431A9B5D3211BE CRC64;
Query Match 6.0%; Score 167; DB 1; Length 153;
Best Local Similarity 30.1%; Pred. No. 6.4e-06;
Matches 43; Conservative 26; Mismatches 48; Indels 26; Gaps 5;
225 VHGTFIDKQVGGD-LEVSYHLBGFNVSDHKHLDHYQYANGDLTSCDNLGAKYDPH 283

```

```

Db 13 VSGVKKFEQASESEPTYSYEIAGNSPNAERGHIEF-----GATNGCVSAGHFHMF 67
Qy 284 EHYHS-----ELCDGDHDDHGVNESHRYSWINIFGDSVGSIAHQ-RDLH 335
Db 68 KTHKAPDPDEVRHNDKMNKVTENGVAKGFSDSLIKLIGPVSYSVIHAGDDLG 127
Qy 336 K-----SAKIACCVYIG 346
Db 128 KDTRESLKTNAGPRACGVIG 150

RESULT 10
SODC_HORSE STANDARD; PRT; 153 AA.
ID SODC_HORSE
AC P00443;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Superoxide dismutase [Cu-Zn] (EC 1.15.1.1).
GN SOD1.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97080551; PubMed=8921896;
RA Divers T.J., Batt C.A.; Wiedmann M., Mohammed H.O., Cummings J.F.,
RT "Equine motor neuron disease is not linked to Cu/Zn superoxide
RT dismutase mutations: sequence analysis of the equine Cu/Zn superoxide
RT dismutase cDNA."
RT Gene 178:83-88(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99261591; PubMed=10331206;
RA Ishida N., Katayama Y., Sato F., Hasegawa T., Mukoyama H.;
RT "The cDNA sequences of equine antioxidant enzyme genes Cu/Zn-SOD and
RT Mn-SOD, and these expressions in equine tissues."
RT J. Vet. Med. Sci. 61:291-294(1999).
RN [3]
RP SEQUENCE.
RX MEDLINE=82052979; PubMed=7298616;
RA Lerch K., Ammer D.;
RT "Amino acid sequence of copper-zinc superoxide dismutase from horse
RT liver."
RT J. Biol. Chem. 256:11545-11551(1981).
CC -1- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -1- COFACTOR: Copper and zinc.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, U38956; AAC48682.1; -.
DR EMBL, AB001692; BAA76921.1; -.
DR PIR, A00515; DSHCZ.
DR HSSP: P00442; 1CBJ.
DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; sodcu.1.
DR PRINTS: PR00068; CUZNDISMASE.
DR PRODOM: PD000469; SOD_CU_ZN.1.

```

```

DR PROSITE:PS00087; SOD_CU_ZN_1; 1.
DR PROSITE:PS00332; SOD_CU_ZN_2; 1.
KM Oxidoreductase; Copper; Zinc; Acetylation.
FT INIT_MET 0
FT MOD_RES 1 1 ACETYLATION.
FT METAL 46 46 COPPER (BY SIMILARITY).
FT METAL 48 48 COPPER (BY SIMILARITY).
FT METAL 63 63 COPPER AND ZINC (BY SIMILARITY).
FT METAL 71 71 ZINC (BY SIMILARITY).
FT METAL 80 80 ZINC (BY SIMILARITY).
FT METAL 83 83 ZINC (BY SIMILARITY).
FT METAL 120 120 COPPER (BY SIMILARITY).
FT METAL 126 146 BY SIMILARITY.
SQ DISULFID 153 AA; 15940 MW; 202A1900EEFA04A3 CRC64;
Query Match 5.9%; Score 164.5; DB 1; Length 153;
Best Local Similarity 31.0%; Pred. No. 9.9e-06;
Matches 45; Conservative 26; Mismatches 43; Indels 31; Gaps 6;
QY 225 VHGTFDKOVGYGDLEVSYHEGF--NVSDHDKHLADYQVLIYANGDLTSGCNDLGAAYD 281
D 14 VHGVIHFEEQOEGRGPRV---LKGFTESLTKGDHGFHVHEF---GINTGGCTTAAGAHN 65
QY 282 F-----HEDYHSELGDLGDIHDDHNVNESHRYSMINIFGDSVYSIAHOR-DH 333
D 66 PLSKHGKGRDEBRIHVGDLGNTVADENKADVDKDVISLSGHSNIGRTVWVHEKODD 125
QY 334 LHK-----SAKIACCVIG 346
D 126 LKGKGNDESTKTGNAGSRACGVIG 150
RESULT 11
SOD4_MAIZE
ID SOD4_MAIZE STANDARD: PRT: 151 AA.
AC P23345;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Superoxide dismutase [Cu-Zn] 4A (EC 1.15.1.1).
GN SODCC.3 OR SOD4A.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Penticostales; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-90136495; PubMed-2482436;
RA Cannon R.E.; Scandallos J.G.;
RT "Two cDNAs encode two nearly identical Cu/Zn superoxide dismutase
RT proteins in maize."
RL Mol. Gen. Genet. 219:1-8(1989).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Copper and zinc.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
CC -!- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X17564; -. NOT_ANNOTATED_CDS.
DR PIR: S07007; S07007.
DR HSP: P07505; ISRD.
DR MAIZEDB: 47586; -.

```

```

DR InterPro: IP001424; sod_cu_zn.
DR Pfam: PF00080; sodcu.1.
DR PRINTS: PR00068; CUZNDISMTASE.
DR ProDom: PD000469; SOD_CU_ZN.1.
DR PROSITE: PS00087; SOD_CU_ZN.1.1.
DR PROSITE: PS00332; SOD_CU_ZN.2.1.
KW Oxidoreductase; Copper; Zinc; Multigene family.
FT INIT_MET 0
FT METAL 44 44 COPPER (BY SIMILARITY).
FT METAL 46 46 COPPER (BY SIMILARITY).
FT METAL 61 61 COPPER AND ZINC (BY SIMILARITY).
FT METAL 69 69 ZINC (BY SIMILARITY).
FT METAL 78 78 ZINC (BY SIMILARITY).
FT METAL 81 81 ZINC (BY SIMILARITY).
FT METAL 118 118 COPPER (BY SIMILARITY).
FT DISULFID 55 144 BY SIMILARITY.
SQ SEQUENCE 151 AA; 14983 MW; 966226F86C919E58 CRC64;

Query Match 5.9%; Score 164; DB 1; Length 151;
Best Local Similarity 31.7%; Pred. No. 1,1e-05;
Matches 45; Conservative 21; Mismatches 50; Indels 26; Gaps 4;

QY 225 VHGTFPKVGVGVDELVSYHLEGFNVSDHDDHDLHDVQIYANGDLTSGCDNLGAKYD-- 282
Db 13 VKSTIFFTEGCDPTTVTGVSGSL-----PGKHGFHVALDITNGCSGSPHNPAS 66
QY 283 -----HEDFHSELGDLGDIHDDHGVNESHRYSMWIFEDSDVSLGSLAIH-ORDHLAK 336
Db 67 KKHGAPEDEENRHHGDTGNVTAGADVCANINVTDSQIFLTPGNSITGRVAVVHADPPDLGK 126
QY 337 -----SAKIACCVIG 346
Db 127 GGHBLSKSTGNAGRVACGIIG 148

RESULT 12
SOD5_MAIZE
ID SOD5_MAIZE STANDARD; PRT; 151 AA.
AC P23346;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Superoxide dismutase [Cu-Zn] 4AP (EC 1.15.1.1).
GN SODCC.2 OR SODAP.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_Taxid:4577;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90136495; PubMed=2482436;
RA Cannon R.E., Scandalios J.G.;
RT "Two cDNAs encode two nearly identical Cu/Zn superoxide dismutase
RT proteins in maize."
RL Mol. Gen. Genet. 219:1-8(1989).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Copper and zinc.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X17565; CAB57992.1; ALT_SEQ.

```



```

DR PIR: S07008; S07008.
DR HSSP: P07505; 1SRD.
DR MaizeDB: 47586; -.
DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; sodcu.1.
DR PRINTS: PR00068; CUZNDISMTASE.
DR PRODOM: PD000469; SOD_CU_ZN.1.
DR PROSITE: PS00087; SOD_CU_ZN.1.
DR PROSITE: PS00332; SOD_CU_ZN.2.1.
DR Oxidoreductase; Copper; Zinc; Multigene family.
KW INIT_MET 0
FT METAL 44 44 COPPER (BY SIMILARITY).
FT METAL 46 46 COPPER (BY SIMILARITY).
FT METAL 61 61 COPPER AND ZINC (BY SIMILARITY).
FT METAL 69 69 ZINC (BY SIMILARITY).
FT METAL 78 78 ZINC (BY SIMILARITY).
FT METAL 81 81 ZINC (BY SIMILARITY).
FT METAL 118 118 COPPER (BY SIMILARITY).
FT DISULFID 55 144 COPPER (BY SIMILARITY).
SQ SEQUENCE 151 AA; 14939 MW; 9C7E572A6C1AEFD CRC64;

Query Match
Best Local Similarity 5.9%; Score 164; DB 1; Length 151;
Matches 45; Conservative 21; Mismatches 50; Indels 26; Gaps 4;

OY 225 VHGTFKQVGVGDLEVSYLEGFNVSDDKHLDHVOIYANGDITLSCDNLGAKYDP-- 282
DB 13 VKGTIFTEBGDGPVAVTGVSGT-----KPLGHGFHVALDITNGCMSTGPHNPNAS 66
OY 283 -----HEDYHSELGDIDHDDHGVNESHRYSWINIFGDDSVLGSIATIH-ORDHLRK 336
DB 67 KENGAPEDETHNADLGNVTAGDGVANINVTDSQIFLTGPNSTIGRAVVVHADPPDLGR 126
OY 337 -----SARIACCVIG 346
DB 127 GGHLSKSTGNAGRVACGIIIG 148

RESULT 13
SODI_ORYSA STANDARD; PRT; 151 AA.
AC P28756;
DT 01-DEC-1992 (Rel. 24, Created)
DT 15-JUN-2002 (Rel. 24, Last sequence update)
DE Superoxide dismutase [Cu-Zn] 1 (EC 1.15.1.1).
GN SODCC1 OR SODCC.1.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC NCBI_TaxID=4530;
OX NCBI_TaxID=4530;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Seed;
RX MEDLINE=92322961; PubMed=1623183;
RA Sakamoto A., Ohsuga H., Tanaka K.;
RT Nucleotide sequences of two cDNA clones encoding different Cu/Zn-
RT superoxide dismutases expressed in developing rice seed (Oryza sativa
RT plant Mol. Biol. 19:323-327(1992)).
RN (2)
RP SEQUENCE FROM N.A.
RC STRATGEV; Jaihong 67; TISSUE=Shoot;
RA Papp S.M., Huang G.B.;
RT Submitted (XXX-1994) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE OF 1-90 FROM N.A.
RX MEDLINE=95241628; PubMed=7724677;
RA Sakamoto A., Okumura K., Kamakura H., Tanaka K.;
RT Molecular cloning of the gene (SODCC1) that encodes a cytosolic
RT copper/zinc-superoxide dismutase from rice (Oryza sativa L.).;
RN Plant Physiol. 107:651-652(1995).

```

```

CC -1- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -1- COFACTOR: Copper and zinc.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D00999; BA00079.1; -.
DR EMBL: U36320; AA03917.1; -.
DR EMBL: U36320; AA03917.1; -.
DR PIR: S22508; S22508.
DR HSSP: P07505; 1SRD.
DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; sodcu.1.
DR PRINTS: PR00068; CUZNDISMTASE.
DR PRODOM: PD000469; SOD_CU_ZN.1.
DR PROSITE: PS00087; SOD_CU_ZN.1.
DR PROSITE: PS00332; SOD_CU_ZN.2.1.
KW Oxidoreductase; Copper; Zinc; Multigene family.
FT INIT_MET 0
FT METAL 44 44 COPPER (BY SIMILARITY).
FT METAL 46 46 COPPER (BY SIMILARITY).
FT METAL 61 61 COPPER AND ZINC (BY SIMILARITY).
FT METAL 69 69 ZINC (BY SIMILARITY).
FT METAL 78 78 ZINC (BY SIMILARITY).
FT METAL 81 81 ZINC (BY SIMILARITY).
FT METAL 118 118 COPPER (BY SIMILARITY).
FT DISULFID 55 144 COPPER (BY SIMILARITY).
FT CONFLICT 56 144 M -> I (1N REF. 2).
SQ SEQUENCE 151 AA; 15120 MW; CF5FD97ACD4E7998 CRC64;

Query Match
Best Local Similarity 5.8%; Score 163; DB 1; Length 151;
Matches 46; Conservative 19; Mismatches 51; Indels 26; Gaps 4;

OY 225 VHGTFKQVGVGDLEVSYLEGFNVSDDKHLDHVOIYANGDITLSCDNLGAKYDP-- 282
DB 13 VKGTIFTEBGDGPVAVTGVSGT-----KPLGHGFHVALDITNGCMSTGPHNPNAG 66
OY 283 -----HEDYHSELGDIDHDDHGVNESHRYSWINIFGDDSVLGSIATIH-ORDHLRK 336
DB 67 KENGAPEDETHNADLGNVTAGDGVANINVTDSQIFLTGPNSTIGRAVVVHADPPDLGR 126
OY 337 -----SARIACCVIG 346
DB 127 GGHLSKSTGNAGRVACGIIIG 148

RESULT 14
SODC_ZANAE STANDARD; PRT; 152 AA.
AC O65174;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Superoxide dismutase [Cu-Zn] (EC 1.15.1.1).
GN SODCC OR SOD3.
OS Zantedeschia aethiopica (White calla lily).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Araceae; Zantedeschia.
OX NCBI_TaxID=69721;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;

```

RA Lino-Neto T., Tavares R.M., Palme K., Pais M.S.S.;
RT "Expression of cytosolic Cu/Zn-superoxide dismutase during senescence
and regression of senile plaques in the hippocampus of Alzheimer's disease
patients (MAR-1998) to the EMBL/GenBank/DBJ databases."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Destroys toxic radicals which are normally produced within the
cells and which are toxic to biological systems.
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -1- COFACTOR: Copper and zinc (By similarity).
CC -1- SUBUNIT: HOMODIMER (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF054150; AAC08581.1; -;
DR HSSP: P07505; ISRD.
DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; sodcu; 1.
DR PRINTS: PR00068; CUZNDISMTASE.
DR ProDom: PD000469; SOD_CU_ZN.1.
DR PROSITE: PS00087; SOD_CU_ZN.1; 1.
DR PROSITE: PS00332; SOD_CU_ZN.2; 1.
KW Oxidoreductase; Copper; zinc.
FT METAL 45 45 COPPER (BY SIMILARITY).
FT METAL 47 47 COPPER (BY SIMILARITY).
FT METAL 62 62 COPPER AND ZINC (BY SIMILARITY).
FT METAL 70 70 ZINC (BY SIMILARITY).
FT METAL 79 79 ZINC (BY SIMILARITY).
FT METAL 82 82 ZINC (BY SIMILARITY).
FT METAL 119 119 COPPER (BY SIMILARITY).
FT METAL 145 145 BY SIMILARITY.
FT DISULFID 56 145
SQ SEQUENCE 152 AA; 15152 MW; 96C535E190102160 CRC64;
Query Match 5.8%; Score 163; DB 1; Length 152;
Best Local Similarity 31.7%; Pred. No. 1.3e-05;
Matches 45; Conservative 22; Mismatches 49; Indels 26; Gaps 4;
QY 225 VHGITDVKQVGYGDEVSYLEGFNSDDKHLHDVQIYANGDLTSGCDNLGAKYDP-- 282
DB 14 VQGVTFEFGQEGGPTTIGSLGL-----KPLGHEFHVALGDTTNGCMSTGPHFNPAG 67
QY 283 -----HEDYHSELGDLGIDHDDHGVVESHRYSWINIFGDDSVLGRSTAIH-QRDHLAK 336
DB 68 KKGAPEDGNRRHAGDLGVTVGDEGTNFTVTDSPILTGLNSVGRVAVHADSDDLGK 127
QY 337 -----SAKIACCVIG 346
DB 128 GGHELSKITGNAGRLACGVIG 149
RESULT 15
SODC_CANAL STANDARD; PRT; 153 AA.
AC 059924;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Superoxide dismutase [Cu-Zn] (EC 1.15.1.1).
GN SOD1.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99233600; PubMed=10216241;
RA Hwang C.S., Rhee G., Kim S.T., Kim Y.R., Huh W.K., Baek Y.U.,

RA Kang S.O.;
RT "Copper- and zinc-containing superoxide dismutase and its gene from
Candida albicans."
RL Biochim. Biophys. Acta 1427:245-255(1999).
CC -1- FUNCTION: Destroys radicals which are normally produced within the
cells and which are toxic to biological systems.
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -1- COFACTOR: Copper and zinc (By similarity).
CC -1- SUBUNIT: HOMODIMER (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF046872; AAC12872.1; -;
DR HSSP: P00445; IUCV.
DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; sodcu; 1.
DR PRINTS: PR00068; CUZNDISMTASE.
DR ProDom: PD000469; SOD_CU_ZN.1.
DR PROSITE: PS00087; SOD_CU_ZN.1; 1.
DR PROSITE: PS00332; SOD_CU_ZN.2; FALSE_NEG.
KW Oxidoreductase; Copper; zinc.
FT INIT_MET 0 0 BY SIMILARITY.
FT METAL 46 46 COPPER (BY SIMILARITY).
FT METAL 48 48 COPPER (BY SIMILARITY).
FT METAL 63 63 COPPER AND ZINC (BY SIMILARITY).
FT METAL 71 71 ZINC (BY SIMILARITY).
FT METAL 80 80 ZINC (BY SIMILARITY).
FT METAL 83 83 ZINC (BY SIMILARITY).
FT METAL 120 120 COPPER (BY SIMILARITY).
FT METAL 146 146 BY SIMILARITY.
FT DISULFID 57 146
SQ SEQUENCE 153 AA; 16015 MW; 8BC67A2D17DCBEOFCRC64;
Query Match 5.8%; Score 163; DB 1; Length 153;
Best Local Similarity 30.1%; Pred. No. 1.3e-05;
Matches 44; Conservative 23; Mismatches 53; Indels 26; Gaps 4;
QY 225 VHGITDVKQVGYGDEVSYLEGFNSDDKHLHDVQIYANGDLTSGCDNLGAKYDPH 283
DB 13 VQGVTFEFGQEGGPTTIGSLGL-----KPLGHEFHVALGDTTNGCMSTGPHFNPAG 67
QY 284 -----EDYHSELGDLGIDHDDHGVVESHRYSWINIFGDDSVLGRSTAIH----- 329
DB 68 KKGAPEDDERHAGDLGNSTGDKNGVAKGTQKODLLIKLKDSILGRTIVHAGTDYDG 127
QY 330 -----QRDHLKSAKIACCVYGRG 349
DB 128 KKGFEDESKITGNAGRLACGVIG 153
Search completed: November 29, 2002, 12:33:14
JOB time : 29 secs


```
Db 21 DGRGNDGKNDKNDHNDHNDHDDDDDETMHYACSEMPRPMASLHHNVHSIEL 80
OY 61 SQRKGAUYLELVLGPNTESEDDHNNHGLHMLGMSKAGCSIGLYANAREKADRG 120
Db 81 SQGHGAUYLELVLGPNTESEDDHNNHGLHMLGMSKAGCSIGLYANAREKADRG 140
OY 121 DLGLDVLDDRGVNVNHYAWLIDGTAPNTEALIGHSMTLLQSGHPTADTPASRIACCV 180
Db 141 DLGLDVLDDRGVNVNHYAWLIDGTAPNTEALIGHSMTLLQSGHPTADTPASRIACCV 200
OY 181 IGHGKARPEAAALHHELEEDKTEHYACDVRSNTHQPKALHHNHVGTIDFKQVGYDLE 240
Db 201 IGHGKARPEAAALHHELEEDKTEHYACDVRSNTHQPKALHHNHVGTIDFKQVGYDLE 260
OY 241 VSYHLEGFNVSDHKDHLHDVQIYANGDLTSGCDNLGAKYDPNHDYHSELGLDIDHDD 300
Db 261 VSYHLEGFNVSDHKDHLHDVQIYANGDLTSGCDNLGAKYDPNHDYHSELGLDIDHDD 320
OY 301 HGYYNESHRSWNTINFGDSDVLSGRSLAHORNDHLKSKAKIACCVIGRGSHPEIYHRAKC 360
Db 321 HGYYNESHRSWNTINFGDSDVLSGRSLAHORNDHLKSKAKIACCVIGRGSHPEIYHRAKC 380
OY 361 VVRPNTESTGLHHNVSGSTFEQTPGGSTHMTADLKGFNVSEDSHNRHGVOLHEMGDS 420
Db 381 VVRPNTESTGLHHNVSGSTFEQTPGGSTHMTADLKGFNVSEDSHNRHGVOLHEMGDS 440
OY 421 HGHSLGRMTHGHDDAHDKPRDGLDVIDSHGIVHSTRTFDHLNVEDLNANSLVIMOG 480
Db 441 HGHSLGRMTHGHDDAHDKPRDGLDVIDSHGIVHSTRTFDHLNVEDLNANSLVIMOG 500
OY 481 GHEVESERVACVIGRA 497
Db 501 GHEVESERVACVIGRA 517

RESULT 2
O9NES7 PRELIMINARY: PRT: 735 AA.
AC O9NES7;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Y38B6.98 protein (Y38B6.p protein).
GN Y38B6.98 or Y38B6A.p.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodermidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN SEQUENCE FROM N.A.
RA Sulston J.E.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA Sulston J.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL132886; CAB60938.1; -
DR EMBL: AL132948; CACS1077.1; -
DR HSP: P13231.1HCE
DR InterPro: IPR002395; Kintogen.
DR PRINTS: PR00934; KINTOGEN.
SEQUENCE 735 AA; 80255 MW; 667B81DBE41159 CRC64;

Query Match 7.38; Score 204.5; DB 5; Length 735;
Best Local Similarity 21.08; Pred. No. 6; 3e-08;
```

```
Matches 102; Conservative 29; Mismatches 178; Indels 177; Gaps 23;
OY 1 DGRGNDGKNDKNDHNDHNDHDDDDDETMHYACSEMPRPMASLHHNVHSIEL 58
Db 386 EGBNHNRAHNDHGVNHNHNGHNGSTNN-----SPAHNGHNGSTNNHNGH 432
OY 59 ELQKGNHGAUYLELVLGPNTESEDDHNNHGLHMLGMSKAGCSIGLYANAREKADRG 110
Db 433 NNARAHNG--HNSHNGHNSPRAHNGHNGHNNHARAHNGHNN-----GENNH 476
OY 111 AREKADRGDLGLYDDRGVNVNHYAWLIDGTAPNTEALIGHSMTLLQSGHPTAD 170
Db 477 A-PANHHNGHNG--TNH-----GH-----HSH---H 498
OY 171 TPASRIACCVIGHKARPEAAALHHELEEDKTEHYACDVRSNTHQPKALHHNHVGTID 224
Db 499 SPAN-----HNGHNNHARAHNGHNGHNGSTNNHNGHNNHARAHNGHNGHNGHNG 550
OY 225 VHGTFDKQVGYDLEVSUYHLEGFNVSDHKDHLHDVQIYANGDLTSGCDNLGAKYDPN 284
Db 551 HNGS-----HNSPRAHNGHNGHNNHARAHNGHNGHNN-----GSHGVNHNHNE 591
OY 285 DY-HSELGLDIDHDDRGVNVNHYAWLIDGTAPNTEALIGHSMTLLQSGHPTAD 343
Db 592 SHGHNHARAHNGHNGHNGHNGHNN-----GAGYGANHNGHNN----- 627
OY 344 VIGRGSHPEIYHRAKCVVRPNTESTGLHHNVSGSTFEQTPGGSTHMTADLKGFNVSED 403
Db 628 -----GANHHNARH-----HNNHNGHNDHNN--GSHGVNHNHNGSTNN-----S 663
OY 404 LSHNRHGVOLHEMGDSHSCCHSLGRMTHGHDDAHDKPRDGLDVIDSHGIVHSTRTFD 463
Db 664 LANHGH-----HGHGTHNGHGH--SPANHNGHGH-----HNGAHNGHNGH 705
OY 464 HLNVED 469
Db 706 HDKEN 711

RESULT 3
O9QZV4 PRELIMINARY: PRT: 726 AA.
AC O9QZV4;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Histidine-rich Ca2+ binding protein.
GN HRC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SKELTAL MUSCLE;
EX MEDLINE=99389148; PubMed=10462052;
RA Ridgeway A.G., Petropoulos H., Siu A., Ball J.K., Skerjanc I.S.;
RT "Cloning, tissue distribution, subcellular localization and
RT overexpression of murine histidine-rich Ca2+ binding protein."
RL FEBS Lett. 456:399-402(1999).
DR EMBL: AF158597; AAD55250.1; -
DR MGD: MGI:96226; HRC.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002049; Laminin_EGF.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS01248; LAMININ_TYPE_EGF; UNKNOWN_1.
SEQUENCE 726 AA; 83623 MW; 6B13F250C5904B30 CRC64;

Query Match 6.4%; Score 178.5; DB 11; Length 726;
Best Local Similarity 21.5%; Pred. No. 7; 6e-06;
Matches 118; Conservative 62; Mismatches 189; Indels 181; Gaps 29;
```

```
Db 117 GHENHNEDLDGSAENHLPQRSHSHEDDGI-----VSSRYHRRVPRNAH 163
QY 56 GSTIELSOKHGCVUYLELHVGFTSEDDHNNHGLHMLGDMAGCDSIGELYNAHPEK 115
Db 164 GHGEEDDDDDGEEER-----RVDMEDSDDNENQVHGH-----QSHSKREDELHNAHSHR 214
QY 116 HADPGELGVLDDRGVYNE-----VNHVAMLDIDGAPRTEALIGHSMTLLGSGTADDT 171
Db 215 HGGHSD-----DDDDGVSTENGQAHNRQDHNEEDDSDSDSHTRVQGRDEHDEDS 270
QY 172 PASRIACCVYIGKARPTAALH-----ELEDKTEYHACDVSNTHQ 217
Db 271 DS-----GEYRHHOTDHQNEEDDDDDDDDEDEKDESTER-----RHQ 311
QY 218 PRALH-----HHV--HGTIDKQVYGDLEVSTHLEGRVNSDQKHLHDVQI 263
Db 312 TQGNKREDEDESDRDHNVSRHG-----RQGTREEDDDDDG--DDSTENHQAHR 363
QY 264 YANGDLTSGCNIGAKYRPHEDYHSELGLDIDHDDHGVVNSHRYSWINIFGDSVYLG 323
Db 364 HRDHEKDED-----DSEEDYH-----HVPSHG--RQSHQ--NEEDEAV- 401
QY 324 RSTAIHQRDHLKSAKIAACVIGSGSHPEIYHRAKCV--RP-----NTSTGLHHNV 375
Db 402 -----STENHQAHRHNAHNLGR--ESEEEVAVKYSHHVAHSHRPOGNADREDSLEEHM 454
QY 376 SGIPTFGPGSTMTADLKGFVNSEDLSHNRHGVQLHENGDMHSGHSLGRMYHGHDD 435
Db 455 N-----EYRGHNNHRAHSGDEDEISTEFGHKAPSHRLQDDE-----RARGHRE 499
QY 436 -----AHDPKRP-----GDLDVYIDDSHGIVHSTFEDHUNVEDLNARSLYIM 478
Db 500 RVQGEIHAHQPLQRTGPSRESRKED-----HSSQSGD-----EDPEQR----- 538
QY 479 OGGHEYESER 488
Db 539 QAHSEEBEEK 548

RESULT 4
QYWE4 PRELIMINARY: PRT: 738 AA.
AC 09WE4;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Histidine-rich calcium-binding protein.
GN HRC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=HEART;
RA Shi S.D., Brunner S.R.;
RT "Molecular cloning of histidine-rich calcium-binding protein from
RT mouse hearts."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF13218; AADA2061.1;
DR MCD; MGI:96225; HRC.
DR InterPro; IPR000561; EGF-like
DR InterPro; IPR002049; Laminin_EGF.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; UNKNOWN_1.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; UNKNOWN_1.
SQ SEQUENCE 738 AA: 85004 MW: D1F6ZDE2D533ECA2 CRC64;

Query Match 6.3%; Score 177; DB 11; Length 738;
Best Local Similarity 21.8%; Pred. No. 1e-05;
Matches 105; Conservative 56; Mismatches 180; Indels 140; Gaps 24;
QY 18 HNDHDDHDDDDDEDTMNY-----AQCEMERPRHMASSLHHNV-----HGSIELSQGHCA 67
```

```
Db 116 HNGHEDNEDLDGSAENHLPQRSHSHEDDGIYSSEYHGVPRNAHGHGEEDDDG 175
QY 68 VYLELHVGFTSEDDHNNHGLHMLGDMAGCDSIGELYNAHPEKADPGDLGVL 127
Db 176 EEEE-----RVDMEDSDDNENQVHGH-----QSHSKREDELHNAHSHRQHS 222
QY 128 DRGVYNE-----VNHVAMLDIDGAPRTEALIGHSMTLLGSGTADDPASRIACCVIGH 183
Db 223 DDDGVSTENGQAHNRQDHNEEDDSDSDSHTRVQGRDEHDEDS 272
QY 184 GARRETAALH-----ELEDKTEYHACDVSNTHQKALH----- 222
Db 273 GEYRHHOTDHQNEEDDDDDDDDEDEKDESTER-----RHQVQGNKREDEDE 323
QY 223 -----HHV--HGTIDKQVYGDLEVSTHLEGRVNSDQKHLHDVQIANGDLTSGCN 275
Db 324 SDDHNVSRHG-----RQGTREEDDDDDG--DDSTENHQAHRHNDHEKDED- 374
QY 276 LQAKYRPHEDYHSELGLDIDHDDHGVVNSHRYSWINIFGDSVYLGSRSTAIHQRDHL 335
Db 375 -----DSEEDYH-----HVPSHG--RQSHQ--NEEDEAV-----STENH 407
QY 336 KSAKIAACVYIGKGSHPREIYHRAKCV--RP-----NTSTGLHHNVSGITFGPFG 387
Db 408 OSPRHARNDLGR--ESEEEVAVKYSHHVAHSHRPOGNADREDSLEEHM--EYVGH 460
QY 388 STMTADLKGFVNSEDLSHNRHGVQLHENGDMHSGHSLGRMYHGHDD-----AHDPK 441
Db 461 HNHRAHSGDEDEISTEFGHKAPSHRLQDDE-----RARGHREPVQGEIHAHQPLQ 511
QY 442 P 442
Db 512 P 512

RESULT 5
QYWE4 PRELIMINARY: PRT: 251 AA.
AC 064466;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Superoxide dismutase (Cu-Zn) (EC 1.15.1.1) (Superoxide dismutase 3,
DE extracellular).
GN SOD3 OR SOD 3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV; TISSUE=LIVER;
RX MEDLINE=95327627; PubMed=7603981;
RA Carlsson L.M., Jonsson J., Edlund T., Marklund S.M.;
RT "Mice lacking extracellular superoxide dismutase are more sensitive to
RT hyperoxia."
RL Proc. Natl. Acad. Sci. U.S.A. 92:6264-6268(1995).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV; TISSUE=LIVER;
RX MEDLINE=96224263; PubMed=8643556;
RA Carlsson L.M., Marklund S.M., Edlund T.;
RT "The rat extracellular superoxide dismutase dimer is converted to a
RT tetramer by the exchange of a single amino acid."
RL Proc. Natl. Acad. Sci. U.S.A. 93:5219-5222(1996).
RN 13
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Neplionev I.V., Folz R.J.;
RT "Mouse Extracellular Superoxide Dismutase (SOD3) gene."
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RN 14
```

RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) -> O(2) + H(2)O(2).
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
DR EMBL; X84940; CAA59335.1; -;
DR EMBL; AF039602; AAC62204.1; -;
DR EMBL; BC010975; AAH10975.1; -;
DR HSSP; P15107; 1XSO.
DR MGD; MGI:103181; Sod3.
DR InterPro; IPR001424; SOD_CU_ZN.
DR Pfam; PF00080; sodcu; 1.
DR PRINTS; PR00068; CUZNDISMUTASE.
DR PRODOM; PD000469; SOD_CU_ZN.1.
DR PROSITE; PS00087; SOD_CU_ZN.1; 1.
DR PROSITE; PS00332; SOD_CU_ZN.2; 1.
KW Copper; Oxidoreductase; Zinc.
SQ SEQUENCE 251 AA; 27409 MW; 27D26FA1610ZCAE CRC64;

Query Match 6.2%; Score 174.5; DB 11; Length 251;
Best Local Similarity 31.1%; Pred. No. 3.8e-06;
Matches 56; Conservative 22; Mismatches 57; Indels 45; Gaps 11;

QY 199 EEDKTEHYAHCDVRSNTQPKAL-----HHVHGTFDFKVGTD-LEVSYHLEGFNVSD 253
DB 67 EVDAEEMHAIQRY-----QPSATLPDPDQPTGLVLFKQGLSGRLSEYFLEGFPAEON 121
QY 254 HKD---HLHDVQIYANGDLTSGCDNLGAKYDPHEDYHSELGLDIDDDHG--VYNES- 307
DB 122 ASNAIVHVEF-----GDSQCGDSTGPRYINMEVPHQ-----HGGFGNYYVANGQ 169
QY 308 ---HRVS-WINIFGDSVLRSTAIHO--RDHLKSA-----KIACCVIGRGS 350
DB 170 LMRHVRGLTSLAGPHSHIGRSVYVAGEDLGGKGNQASLQNGNAGRRLACCVGTSS 229

RESULT 6
Q9ZNO4 PRELIMINARY; PRT; 152 AA.
AC Q9ZNO4;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Superoxide dismutase (EC 1.15.1.1) (CU-ZN).
GN SODCC OR SOD-1.
OS Cloner arletium (Chickpea) (Garbanzo).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Euphorbia; Euphorbiaceae; Euphorbiaceae; Euphorbiaceae; Clitorea; Clitorea.
OX NCBI_TaxID=827;
RA SEQUENCE FROM N.A.
RC STRAIN-CV. IL3279; TISSUE=LEAF;
RA Hanselle T., Barz W.;
RT "Isolation of cDNA Encoding Superoxide Dismutase from Infected
Chickpea Plants."
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. CASTELLANA; TISSUE=ETIOLATED EPICOTYLS;
RA Dopico B., Hernandez-Nistal J., Labrador E.;
RT "A cDNA encoding a superoxide dismutase is expressed in chickpea
epicotyls."
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -1- CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) -> O(2) + H(2)O(2).
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
DR EMBL; A010239; CA10160.1; -;
DR EMBL; A010261; CA10132.1; -;
DR HSSP; P07505; ISRD.

DR InterPro; IPR001424; SOD_CU_ZN.
DR Pfam; PF00080; sodcu; 1.
DR PRINTS; PR00068; CUZNDISMUTASE.
DR PRODOM; PD000469; SOD_CU_ZN.1.
DR PROSITE; PS00087; SOD_CU_ZN.1; 1.
DR PROSITE; PS00332; SOD_CU_ZN.2; 1.
KW Copper; Oxidoreductase; Zinc.
SQ SEQUENCE 152 AA; 15222 MW; 4ED0AF0A54A8A54 CRC64;

Query Match 6.2%; Score 174; DB 10; Length 152;
Best Local Similarity 33.1%; Pred. No. 2.1e-06;
Matches 47; Conservative 23; Mismatches 46; Indels 26; Gaps 4;

QY 225 VHGTFKQVGYGDLVSYHLEGFNVSDHDKHLDVQIYANGDLTSGCDNLGAKYDPH- 283
DB 14 VSGTINSGQGDGPTVTGNTLGL-----KPLGHGFHIALGDTTNGCISTGPHNPNMG 67
QY 284 -----EDYHSELGLDIDDDHGCVNESHRYSWINIFGDSVLRSTAIH-QRDLHK 336
DB 68 KEGSPEDPIRHAAGDLGNINVGDSGTVSFSITDNLPLTGNSTIGRAVYVHADPDGLCK 127
QY 337 -----SAKIACCVIG 346
DB 128 GGHLEKSTTGNAGRVACGIG 149

RESULT 7
Q90831 PRELIMINARY; PRT; 1840 AA.
AC Q90831;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Translatin.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RA SEQUENCE FROM N.A.
RC MEDLINE-9530324; PubMed-7781165;
RA Kelly M.M., Phanthourath C., Brees D., McCabe C.F., Cole G.J.;
RT "Molecular characterization of RAP-300, a high molecular weight
embryonic polypeptide containing an amino acid repeat comprised of
multiple leucine-zipper motifs."
RL Brain Res. Dev. Brain Res. 85:31-47(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC CODE G.J.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; X80877; CAA56845.1; -;
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; Efilament; 1.
DR PROSITE; PS00226; IF; UNKNOWN_1.
SQ SEQUENCE 1840 AA; 202279 MW; 8ED355F3F474831 CRC64;

Query Match 6.2%; Score 173; DB 13; Length 1840;
Best Local Similarity 22.7%; Pred. No. 7.3e-05;
Matches 111; Conservative 60; Mismatches 175; Indels 142; Gaps 32;

QY 11 KDHNND--DHNDDNDDDDDDDETGNVACSEMERPNHASSLNHNHVGSTEL-----SQGK 64
DB 825 REEHRLDQEGHGDGLQVNEHDLWEQGDIOEE-----HGDKQEVGTQEE 869
QY 65 HGAUVLELHNVGNTSEDDND--HNHGLHLNMGMSACSGTIGVNAPEKNAPRGL 122
DB 870 HG-----DLQEGSGDLQENHGDQVNEHDLWEQGDIOEE-----EENHGDVNEHDLQVHGL 922
QY 123 -----GDL-----VDHDSGVNVENHNAW-----LDLQAPVTEALHNSMTLQSGHTDA 169
DB 923 QEHNGDLQVNEHDLQVNEHDLQVNEHDLQVNEHDLQVNEHDLQVNEHDLQVNEHDLQV 978


```

RL Genetics 144:317-328(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-90136495; PubMed-2482436;
RA Cannon R.E., Scandalios J.G.;
RT "Two cDNAs encode two nearly identical Cu/Zn superoxide dismutase
   proteins in maize."
RL Mol. Gen. Genet. 219:1-8(1989).
CC -!- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) -> O(2) + H(2)O(2).
CC -!- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
DR EMBL: U34727; AAB49913.1; -.
DR EMBL: X17564; CAB57993.1; -.
DR HSSP: P07505; 1SRD.
DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; sodcu.1.
DR PRINTS: PR00068; CUZNDISMUTASE.
DR PRODOM: PD000469; SOD_CU_ZN.1.
DR PROSITE: PS00087; SOD_CU_ZN.1.
DR PROSITE: PS00332; SOD_CU_ZN.2; 1.
KM COPPER: Oxidoreductase; Zinc.
SQ SEQUENCE 152 AA; 15089 MW; 691DF62F88A7623 CRC64;

Query Match 5.8%; Score 163; DB 10; Length 152;
Best Local Similarity 31.7%; Pred. No. 1.6e-05;
Matches 45; Conservative 21; Mismatches 50; Indels 26; Gaps 4;

OY 225 VHGTFDFKQVGGDLEVSYLEGFVNSDDHDKHLHDVOIYANGDLISGCDNLGAKYDP-- 282
DB 14 VKCTIFFDGDDGPTAVNTGVSGL-----KPGHLGHFVHALGDTNGCKSTGHDYNPAS 67
OY 283 -----HEDYHSELGDIHDHDDHGVNESHRYSMINIFGDDSVLGRSAIR-ORDHLK 336
DB 68 KKHGKDEDENRHAGDIGNVTAGADANVINNTDSQIPLTGNSIIRAVVYHADPDLDLK 127
OY 337 -----SAKIACCVIG 346
DB 128 GGHLSKSTGNAGRVACGILG 149

RESULT 13
O8WNN6 PRELIMINARY; PRT; 153 AA.
ID O8WNN6;
AC O8WNN6;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Cu/Zn superoxide dismutase.
GN SOD1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Green S.L., Tolwani R.J., Varma S., Quignon P., Gallibert F.,
RA Cork L.C.;
RT "The canine Cu/Zn superoxide dismutase gene (SOD1).";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF346417; AAL61608.1; -.
DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; sodcu.1.
DR PRINTS: PR00068; CUZNDISMUTASE.
DR PRODOM: PD000469; SOD_CU_ZN.1.
DR PROSITE: PS00332; SOD_CU_ZN.2; 1.
SQ SEQUENCE 153 AA; 15912 MW; 0D7900E59C57E6H0 CRC64;

Query Match 5.8%; Score 163; DB 6; Length 153;
Best Local Similarity 31.7%; Pred. No. 1.6e-05;
Matches 46; Conservative 22; Mismatches 51; Indels 26; Gaps 5;

OY 225 VHGTFDFKQVGGDLEVSYLEGFVNSDDHDKHLHDVOIYANGDLISGCDNLGAKYDP-- 282

```

```

DB 15 VECTHFFVKGSGPVVSGTITGL-TEGEHGFHGFPE-----DXTGCTSGAPHFNPIS 68
OY 283 -----HEDYHSELGDIHDHDDHGVNESHRYSMINIFGDDSVLGRSAIRH-ODHLK 336
DB 69 KKHGKDEDENRHAGDIGNVTAGADANVINNTDSQIPLTGNSIIRAVVYHADPDLDLK 128
OY 337 -----SAKIACCVIGRQ 349
DB 129 GDNESTGTGNAGSLACGVLIGIQ 153

RESULT 14
O9V523 PRELIMINARY; PRT; 179 AA.
ID O9V523;
AC O9V523;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE CG9027 protein.
GN CG9027.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Ephyroptera; Neoptera; Emdopterygota; Diptera; Brachycera; Muscomorpha;
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutcliffe G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostel C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Mei K.-H., Idegami C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshireli A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacelb J.M.,
RA Palazozo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Spidlen Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spindling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhan G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) -> O(2) + H(2)O(2).
CC -!- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
DR EMBL: AEO03826; AAF58647.2; -.

```

```

DR HSP; P07505; 1SRD.
DR FLYBase; FBgn0033631; CG9027.
DR InterPro; IPR001424; SOD_CU_ZN.
DR Pfam; PF00080; sodcu.1.
DR PRINTS; PR00068; CUZNDISMTASE.
DR PRODOM; PD000469; SOD_CU_ZN.1.
DR PROSITE; PS00087; SOD_CU_ZN.1.
DR PROSITE; PS00332; SOD_CU_ZN.2; 1.
DR Copper; Oxidoreductase; Zinc.
SQ SEQUENCE 179 AA; 18950 MW; 8C45D94E22D81B9F CRC64;

Query Match 5.8%; Score 162.5; DB 5; Length 179;
Best Local Similarity 27.6%; Pred. No. 2.2e-05;
Matches 53; Conservative 31; Mismatches 61; Indels 47; Gaps 9;

OY 333 HLHSAKACVIGRGQSHP--EIVHRAKCVYRPMTSTGILHHVSGSTPEOTP--GGST 389
DB 4 YLVVSLALCATICSAQTNPICAIAYLIGPVQSDNTQ-----VKGNVFTQNDCCGV 56
OY 390 HMTADLGFVNSIEDLSHHRGVGLHMGMSHSGHSLGMY-----HGDDAHDPKRP 443
DB 57 HVAVQLG-----LKEKGKGFHHEKGLTNGCISMGAHNPKVDGCPD--HEVAVHG 109
OY 444 DLGVINDSHGTVHSTPDPHNV--DLNARSIVMGCHVESE----- 487
DB 110 DLGNLEANSIGIIVTDDVITLTGKLGITGRGVV-----HELEDGLGNGHDSKKT 165
OY 488 ----RVACVYG 495
DB 166 NAGGRIACGVIG 177

RESULT 15
OY17FO PRELIMINARY; PRT; 181 AA.
ID OY17FO;
AC 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE CG9027 protein.
GN CG9027.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Ephydroidea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC NCBI_TaxID=7227;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanalides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Butler H., Brokstein P., Brothier P.,
RA Burlis K.C., Busam D.A., Butler K., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Delcher B., Deng Z., Deyanport L.B., Davies P.,
RA de Pablos B., Delcher S., Dahlke C., Davenport L.B., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.G., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.H., Idegawa C.,
RA Jaislin M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel P.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laesko B., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

```

```

RA Merklov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy I., Murry D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Pui V., Reese M.G.,
RA Relnert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Slier E.C., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wattaman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhou X., Zhu S., Zhu X.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu G., Zhao Q., Zheng L.,
RA Gjbbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) -> O(2) + H(2)O(2).
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
DR HSP; P07505; 1SRD.
DR FLYBase; FBgn0033631; CG9027.
DR InterPro; IPR001424; SOD_CU_ZN.
DR Pfam; PF00080; sodcu.1.
DR PRINTS; PR00068; CUZNDISMTASE.
DR PRODOM; PD000469; SOD_CU_ZN.1.
DR PROSITE; PS00087; SOD_CU_ZN.1.
DR PROSITE; PS00332; SOD_CU_ZN.2; 1.
DR Copper; Oxidoreductase; Zinc.
SQ SEQUENCE 181 AA; 19221 MW; 68EACE43F8D05CA1 CRC64;

Query Match 5.8%; Score 162.5; DB 5; Length 181;
Best Local Similarity 31.0%; Pred. No. 2.2e-05;
Matches 45; Conservative 25; Mismatches 44; Indels 31; Gaps 7;

OY 225 VHGTFDKROYG-V-DLEVSHLEGFNVSDDHDKHLDHVDIYANGDLTSGCNDLAKYDP- 282
DB 43 VKGNVFTQNDCCGVHVRVQLG-----KEKGKGFHHEKGLTNGCISMGAHNPKVD 96
OY 283 -----HEDYHSELGLDIDHDDHGVNESHRYSWINIFGDSVLSGSIATHO-RDH 333
DB 97 KVDHGGDPDHVRH--VGDLGNLEANSIGIIVTDDVITLTGKLGITGRGVVHELED 154
OY 334 L---HKSAR-----IACVYG 346
DB 155 LGLGNGHDSKKTGNAGRIACGVIG 179

```

Search completed: November 29, 2002, 12:35:07
 Job time : 95 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2002, 12:37:00 ; Search time 33 Seconds
(without alignments)
2006.835 Million cell updates/sec

Title:	US-09-868-760-7
Perfect score:	497
Sequence:	1 DGEQCNDGQNKDDHDDHDD.....MGGCHEVESERVACVIGRA 497

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

word size :

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Listing first 45 summaries

```
Database :
A: geneseq_101002.*
1: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/genseq/genseqp-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/genseq/genseqp-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/genseq/genseqp-emb1/AA2002.DAT.*
```

SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
1	497	100.0	497	21	AAV93750		Amino acid sequence
19	19.	3.8	19	21	AAV93745		peptide derived fr
3	10	2.0	10	21	AAV93745		peptide derived fr
4	9	1.8	9	21	AAV93747		peptide derived fr
5	8	1.6	8	22	ABR41196		peptide derived fr
6	8	1.6	18	22	ABR25212		peptide #8702 enco
7	8	1.6	18	22	AAAG62056		Protein #7211 enco
8	8	1.6	18	22	AAAG62056		Human brain expres
9	8	1.6	18	22	AAAG74858		Human bone marrow
9	8	1.6	18	22	AAAG34975		peptide #9012 enco
10	8	1.6	18	23	ABG44634		Human peptide enco

1	Arabidopsis thaliana	1.6	122	21	AAG38770
2	Arabidopsis thaliana	1.6	122	21	AAG38770
3	Arabidopsis thaliana	1.6	122	21	AAG38770
4	Arabidopsis thaliana	1.6	125	21	AAG08769
5	Arabidopsis thaliana	1.6	125	21	AAG38768
6	Arabidopsis thaliana	1.6	259	22	AU054288
7	Arabidopsis thaliana	1.6	474	6	AAU50587
8	A. nidulans ribulose	1.6	476	6	AAU50584
9	Tabacco ribulose-b	1.6	477	6	AAU50586
10	Splachn ribulose-b	1.6	1445	20	AAU93595
11	COR. potestium thapa	1.6	1529	17	AAU97985
12	O. longistaminata	1.6	1529	17	AAU93595
13	Peptide derived from	1.4	7	21	AAU93748
14	Peptide derived from	1.4	7	21	AAU93749
15	Novel human diaphn	1.4	34	22	ABG16377
16	Human ORF469 prote	1.4	58	23	ABP13496
17	Proiononbacterium	1.4	66	22	AAU60940
18	Proiononbacterium	1.4	108	22	AAU54154
19	Human protein isol	1.4	125	23	ABH72293
20	Proiononbacterium	1.4	127	22	AAU63370
21	Arabidopsis thaliana	1.4	172	21	AAG35153
22	Arabidopsis thaliana	1.4	172	21	AAG35156
23	Novel signal trans	1.4	192	22	AAU17298
24	Proiononbacterium	1.4	205	22	AAU44674
25	Human immune/thema	1.4	222	22	AAH83152
26	Novel signal trans	1.4	222	22	AAU17610
27	Arabidopsis thaliana	1.4	325	21	AAG35152
28	Arabidopsis thaliana	1.4	326	21	AAG28104
29	Arabidopsis thaliana	1.4	326	21	AAG35155
30	Human secreted prote	1.4	247	22	AAH86931
31	Human secreted prote	1.4	247	22	AAH86931
32	Human secreted prote	1.4	253	19	AAW74976
33	Arabidopsis thaliana	1.4	253	21	AAG08785
34	Lactococcus lactis	1.4	259	23	ABBS5400
35	Arabidopsis thaliana	1.4	260	21	AAG442023
36	Arabidopsis thaliana	1.4	268	23	AAU935084
37	Lactococcus lactis	1.4	275	23	ABBS5512

ALIGNMENTS

RESULT	1
AA	AY93750
ID	AY93750 standard; Protein; 497 AA.

AC	AAV93750;	
XX		
DT	03-OCT-2000 (first entry)	
XX		
DE	Amino acid sequence of a serine protease inhibitor protein.	
XX		
KM	Serine protease inhibitor; green-lipped mussel; anti-thrombin	
KW	divalent metal cation binding activity; dietary supplement.	
KW	anticoagulant.	

PT anti-thrombin activity and divalent metal cation binding activity,
 PT useful as an anticoagulant agent and as a dietary supplement -
 PS Claim 2; Page 2; 44pp; English.
 XX
 CC The present sequence represents a serine protease inhibitor
 CC protein. The protein is isolated from the green-lipped mussel
 CC (Perna canaliculus), and exhibits, inter alia, anti-thrombin activity
 CC and divalent metal cation binding activity. The serine protease
 CC inhibitor protein has a molecular weight of about 55 kilo Daltons.
 CC The protein, and its fragments, are useful in medicaments, in food,
 CC as dietary supplements or as bioremediation agents. In the dietary
 CC supplements, the protein is associated with or bound to at least one
 CC divalent cation (such as calcium, magnesium or zinc) of dietary
 CC significance. The proteins or their fragments are also useful as
 CC anticoagulant agents.
 CC
 SQ Sequence 497 AA:

Query Match 100.0%; Score 497; DB 21; Length 497;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DGEQNDGKNDKDDHNDHDDDDDETMHYACSEMPNPMASLHHVHGSIEL 60
 Db 1 DGEQNDGKNDKDDHNDHDDDDDETMHYACSEMPNPMASLHHVHGSIEL 60
 OY 61 SOKHGAVYLELHVGFTSDEHDHNGHGLHMLGDSAGCSIGELYNAHPEKHADPG 120
 Db 61 SOKHGAVYLELHVGFTSDEHDHNGHGLHMLGDSAGCSIGELYNAHPEKHADPG 120
 OY 121 DLGDLVDVDRGVNENHYAMLDIDTAPNTEALIGHSMITLLOGSTHDAFTPSRIACCV 180
 Db 121 DLGDLVDVDRGVNENHYAMLDIDTAPNTEALIGHSMITLLOGSTHDAFTPSRIACCV 180
 OY 181 IGHGKARPEETAALHNELEEDKTEHYAHCDVRSNTQPKALHHNVGTIDFQOVGYDLE 240
 Db 181 IGHGKARPEETAALHNELEEDKTEHYAHCDVRSNTQPKALHHNVGTIDFQOVGYDLE 240
 OY 241 VSYHLEGFVNSDDKHDLADVOIYANGDLTSGCDNLGAKTPREDHNSFGLDGDHDD 300
 Db 241 VSYHLEGFVNSDDKHDLADVOIYANGDLTSGCDNLGAKTPREDHNSFGLDGDHDD 300
 OY 301 HGVNESHHSYMINIFGDDSVLGSIATHORDHAKSAKTAACCVIGRQSHPEIYHAKCV 360
 Db 301 HGVNESHHSYMINIFGDDSVLGSIATHORDHAKSAKTAACCVIGRQSHPEIYHAKCV 360
 OY 361 VVAPRTSTGLHHVSGSITFEQTPGSGTHAKTADLKGFNVSEDLSHHRRGVOLHEWGDMS 420
 Db 361 VVAPRTSTGLHHVSGSITFEQTPGSGTHAKTADLKGFNVSEDLSHHRRGVOLHEWGDMS 420
 OY 421 HGHSTLGRMYHGHDDAHDPKRRGDLGVDIDSHGIVHSTFPHLNVEDLNARSLVIMOG 480
 Db 421 HGHSTLGRMYHGHDDAHDPKRRGDLGVDIDSHGIVHSTFPHLNVEDLNARSLVIMOG 480
 OY 481 GHEVESERVACVIGRA 497
 Db 481 GHEVESERVACVIGRA 497

RESULT 2
 ID AAY93746 standard; peptide: 19 AA.
 AC AAY93746;
 XX
 DT 03-OCT-2000 (first entry)
 XX
 DE Peptide derived from a serine protease inhibitor protein.
 XX
 KW Serine protease inhibitor; green-lipped mussel; anti-thrombin;
 KW divalent metal cation binding activity; dietary supplement;
 KW anticoagulant.

XX Perna canaliculus.
 OS
 XX WO200039165-A1.
 PN
 XX
 PD 06-JUL-2000.
 XX
 PD 23-DEC-1999; 99WO-N200227.
 PF
 XX 23-DEC-1998; 98NZ-0333568.
 PR 23-JUL-1999; 99NZ-0336906.
 XX
 PA (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.
 XX
 PI Scotti PD, Dearing SC, Greenwood DR, Newcomb RD;
 XX
 DR WPI: 2000-452375/39.
 XX

PT New Perna canaliculus serine protease inhibitor protein exhibiting
 PT anti-thrombin activity and divalent metal cation binding activity,
 PT useful as an anticoagulant agent and as a dietary supplement -
 PS Claim 1; Page 16; 44pp; English.

AA93745-49 represent peptides derived from a serine protease inhibitor
 CC protein. The protein is isolated from the green-lipped mussel
 CC (Perna canaliculus), and exhibits, inter alia, anti-thrombin activity
 CC and divalent metal cation binding activity. The serine protease
 CC inhibitor protein has a molecular weight of about 55 kilo Daltons.
 CC The protein, and its fragments, are useful in medicaments, in food,
 CC as dietary supplements or as bioremediation agents. In the dietary
 CC supplements, the protein is associated with or bound to at least one
 CC divalent cation (such as calcium, magnesium or zinc) of dietary
 CC significance. The proteins or their fragments are also useful as
 CC anticoagulant agents.
 CC
 SQ Sequence 19 AA:

Query Match 3.8%; Score 19; DB 21; Length 19;
 Best Local Similarity 100.0%; Pred. No. 1,8e-11;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 479 OSGHEVESERVACVIGRA 497
 Db 1 OSGHEVESERVACVIGRA 19

RESULT 3
 ID AAY93745 standard; peptide: 10 AA.
 AC AAY93745;
 XX

DT 03-OCT-2000 (first entry)
 XX

DE Peptide derived from a serine protease inhibitor protein.

XX Serine protease inhibitor; green-lipped mussel; anti-thrombin;
 KW divalent metal cation binding activity; dietary supplement;
 KW anticoagulant.

OS Perna canaliculus.
 XX
 PN WO200039165-A1.
 XX

PD 06-JUL-2000.
 XX

PF 23-DEC-1999; 99WO-N200227.
 XX

PR 23-DEC-1998; 98NZ-0333568.
 XX

PR 23-JUL-1999; 99NZ-0336906.
 XX

PA (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.

PI Scottl PD, Dearing SC, Greenwood DR, Newcomb RD;
XX
DR WPI; 2000-452375/39.
XX
PT New Perna canaliculus serine protease inhibitor protein exhibiting
PT anti-thrombin activity and divalent metal cation binding activity,
XX useful as an anticoagulant agent and as a dietary supplement -
PS Claim 1; Page 16; 4app; English.

CC AAY93745-49 represent peptides derived from a serine protease inhibitor
CC protein. The protein is isolated from the green-lipped mussel
CC (Perna canaliculus), and exhibits, inter alia, anti-thrombin activity
CC and divalent metal cation binding activity. The serine protease
CC inhibitor protein has a molecular weight of about 55 kilo Daltons.
CC The protein, and its fragments, are useful in medicaments, in food,
CC as dietary supplements or as bioremediation agents. In the dietary
CC supplements, the protein is associated with or bound to at least one
CC divalent cation (such as calcium, magnesium or zinc) of dietary
CC significance. The proteins or their fragments are also useful as
CC anticoagulant agents.

SQ Sequence 10 AA:

Query Match 2.0%; Score 10; DB 21; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DGEECNDGON 10
| | | | | | | | | |
Db 1 DGEECNDGON 10

RESULT 4
ID AAY93747 standard; peptide: 9 AA.
AC AAY93747:
DT 03-OCT-2000 (first entry)
XX
DE Peptide derived from a serine protease inhibitor protein.
XX
RW Serine protease inhibitor; green-lipped mussel; anti-thrombin;
KW divalent metal cation binding activity; dietary supplement;
RM anticoagulant.
XX
OS Perna canaliculus.
PN WO200039165-A1.
XX
PD 06-JUL-2000.
XX
PF 23-DEC-1999; 99WO-NZ00227.
XX
PR 23-DEC-1998; 98NZ-0333568.
PR 23-JUL-1999; 99NZ-0336906.
XX
PA (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.
XX
PI Scottl PD, Dearing SC, Greenwood DR, Newcomb RD;
XX
DR WPI; 2000-452375/39.
XX
PT New Perna canaliculus serine protease inhibitor protein exhibiting
PT anti-thrombin activity and divalent metal cation binding activity,
XX useful as an anticoagulant agent and as a dietary supplement -
XX Claim 1; Page 16; 4app; English.

CC AAY93745-49 represent peptides derived from a serine protease inhibitor
CC protein. The protein is isolated from the green-lipped mussel

		(Pernis canaliculus), and exhibits inter alia, anti-thrombin activity
CC		and divalent metal cation binding activity. The serine protease
CC		inhibitor protein has a molecular weight of about 55 Kilo Daltons.
CC		The protein, and its fragments, are useful in medicaments, in food,
CC		as dietary supplements or as bioremediation agents. In the dietary
CC		supplements, the protein is associated with or bound to at least one
CC		divalent cation (such as calcium, magnesium or zinc) of dietary
CC		significance. The proteins or their fragments are also useful as
XX		anticoagulant agents.
SQ	Sequence	9 AA:
OY	Query Match	1.8%; Score 9; DB 21; Length 9;
	Best Local Similarity	100.0%; Pred. No.7.8e+05;
	Matches 9; Conservative	0; Mismatches 0; Indels 0; Gaps
	348 GOSHPETIVH 356	
	I	
DB	1 GOSHPETIVH 9	
RESULT 5		
ID	ABB41196 standard; Peptide; 18 AA.	
XX	ABB41196	
AC	ABB41196;	
XX		
DT	04-FEB-2002 (first entry)	
DE	Peptide #8702 encoded by human foetal liver single exon probe.	
XX		
KW	Human; foetal liver; gene expression; single exon nucleic acid probe.	
XX		
OS	Homo sapiens.	
XX		
PM	WO200157277-A2.	
PD		
XX	09-AUG-2001.	
XX		
PF	30-JAN-2001; 2001MO-US00669.	
XX		
PR	04-FEB-2000; 2000US-0180312.	
PR	26-MAY-2000; 2000US-0207456.	
PR	30-JUN-2000; 2000US-0608408.	
PR	03-AUG-2000; 2000US-0632667.	
PR	21-SEP-2000; 2000US-0234687.	
PR	27-SEP-2000; 2000US-0236359.	
PR	04-OCT-2000; 2000GB-0024263.	
XX		
PA	(MOL-) MOLECULAR DYNAMICS INC.	
XX		
PI	Penn SG, Hanzel DK, Chen W, Rank DR;	
XX		
DR	WPI; 2001-483447/52.	
XX		
PT	Human genome-derived single exon nucleic acid probes useful for	
XX	analyzing gene expression in human fetal liver -	
PS		
Claim 27; SEQ ID NO 33831; 639pp + sequence listing; English.		
XX		
CC	The invention relates to a single exon nucleic acid probe for	
CC	measuring human gene expression in a sample derived from human foetal	
CC	liver. The single exon nucleic acid probes may be used for predicting,	
CC	measuring and displaying gene expression in samples derived from human	
CC	fetal liver. The present sequence is a peptide encoded by a single exon	
CC	nucleic acid probe of the invention.	
CC	Note: The sequence data for this patent did not form part of the	
CC	printed specification, but was obtained in electronic format directly	
CC	from WIPO at ftp.wipo.int/pub/published_pat_sequences.	
XX		
SQ	Sequence	18 AA:

Query Match 1.6%; Score 8; DB 22; Length 18;

Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 23 HDDHDDD 30
 |||||
DB 9 HDDHDDD 16

RESULT 6
ABR25212
ID ABR25212 standard; Protein: 18 AA.
XX
AC ABR25212;
XX
DT 23-JAN-2002 (first entry)
XX
DE Protein #7211 encoded by probe for measuring heart cell gene expression.
XX
KW Human: gene expression; heart; microarray; vascular system;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 15; SEQ ID No 26982; 530pp; English.
XX
SQ The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart (see
CC ABR21535-ABR41305). The present sequence is a protein encoded by one such
CC probe. The probes may be used for predicting, measuring and displaying
CC gene expression in samples derived from the human heart via microarrays.
CC By measuring gene expression, the probes are useful for predicting
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease,
CC hypertension, cardiac arrhythmias and congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at http://wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 18 AA:

Query Match 1.6%; Score 8; DB 22; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 23 HDDHDDD 30
 |||||
DB 9 HDDHDDD 16

RESULT 7
AAM62056
ID AAM62056 standard; Protein: 18 AA.
XX
AC AAM62056;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 34161.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 34161; 650pp + Sequence Listing; English.
XX
SQ The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
SQ Sequence 18 AA:

Query Match 1.6%; Score 8; DB 22; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 23 HDDHDDD 30
 |||||
DB 9 HDDHDDD 16

RESULT 8
AAM74858
ID AAM74858 standard; Protein: 18 AA.
XX
AC AAM74858;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 35164.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukemia; lymphoma; myeloma.
XX

```

OS Homo sapiens.
XX
XX WO200157276-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000668.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 35164; 658bp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention.
XX
XX Sequence 18 AA;
XX
XX Query Match 1.6%; Score 8; DB 22; Length 18;
XX Best Local Similarity 100.0%; Pred. No. 1.8;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 23 HDDHDDD 30
XX |||||||
XX Db 9 HDDHDDD 16
XX
XX RESULT 9
XX AAM34975
XX ID AAM34975 standard; Protein; 18 AA.
XX
XX AAM34975;
XX
XX 17-OCT-2001 (first entry)
XX
XX Peptide #9012 encoded by probe for measuring placental gene expression.
XX
XX Probe: microarray; human; placenta; antenatal diagnosis;
XX genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.

```

```

XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-48897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 27; SEQ ID No 35244; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP;
XX see A131315-A157546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders.
XX
XX Sequence 18 AA;
XX
XX Query Match 1.6%; Score 8; DB 22; Length 18;
XX Best Local Similarity 100.0%; Pred. No. 1.8;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 23 HDDHDDD 30
XX |||||||
XX Db 9 HDDHDDD 16
XX
XX RESULT 10
XX ABG44634
XX ID ABG44634 standard; Peptide; 18 AA.
XX
XX ABG44634;
XX
XX 19-AUG-2002 (first entry)
XX
XX Human peptide encoded by genome-derived single exon probe SEQ ID 34299.
XX
XX Human; single exon probe; asthma; lung cancer; COPD; ILD;
XX chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemorrhoidosis;
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Kargener syndrome;
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX primary ciliary dyskinesia; pulmonary hypertension;
XX hyaline membrane disease.
XX
XX Homo sapiens.
XX
XX WO200166003-A2.
XX
XX 15-NOV-2001.
XX
XX 30-JAN-2001; 2001WO-US000665.
XX
XX 04-FEB-2000; 2000US-180312P.
XX
XX 26-MAY-2000; 2000US-207456P.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-234687P.
XX
XX 27-SEP-2000; 2000US-236359P.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2002-114183/15.
XX
XX Spatially-addressable set of single exon nucleic acid probes, used to

```

PT	measure gene expression in human lung samples -
XX	
PS	Claim 27; SEQ ID NO 34299; 634pp; English.

PS Claim 27; SEQ ID No 34299; 634pp; English.

The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human lung comprising single exon nucleic acid probes having one of 12614 nucleic acid sequences mentioned in the specification, or their complements or the 12387 open reading frames derived from the 12614 probes. Also included are a microarray comprising the novel set of probes; the novel set of probes which hybridise at high stringency to a nucleic acid expressed in the human lung, measuring gene expression in a sample derived from human lung, comprising (a) contacting the array with a sample collection of detectably labeled nucleic acids derived from human lung mRNA, and (b) measuring the labeled detectably bound to each probe of the array; identifying exons in a eukaryotic genome, comprising (a) algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exons, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarrays having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary hemangioidosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Kartagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present invention is a peptide/protein encoded by a single exon probe of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIP0 at [ftp.wipo.int/pub/publ/published_pat_sequences](http://wipo.int/pub/publ/published_pat_sequences).

SQ Sequence 18 AA;

Query Match	1.68;	Score 8;	DB 23;	Length 18;
-------------	-------	----------	--------	------------

Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY	23	HDDHDDDD	30
Db	9	HDDHDDDD	16

Db 9 HHHHHHHH 16

RESULT 11	
AAG08770	
ID	AAG08770 standard; Protein; 122 AA

AC AAG08770

DT	17-OCT-2000 (first entry)
----	---------------------------

Accession	Protein	Fragment	Seq ID	NO
DE	Arabidopsis thaliana	protein	Fragment	Seq ID NO: 6439

Protein identification; signal transduction pathway; metabolic pathway;

hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

Arabidopsis thaliana

AA
PN
EP1033405-A2

XX	06-SEP-2000.
PD	
XX	
PF	25-FEB-2000; 2000EP-0301439

[illegible]

PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142903.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	26-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143624.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	07-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	08-OCT-1999;	99US-0158029.
PR	19-JUL-1999;	99US-0144332.	PR	12-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144334.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144334.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159329.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159331.
PR	21-JUL-1999;	99US-0144684.	PR	14-OCT-1999;	99US-0159337.
PR	21-JUL-1999;	99US-0145085.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0145085.	PR	18-OCT-1999;	99US-0159638.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160741.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160768.
PR	23-JUL-1999;	99US-0145142.	PR	21-OCT-1999;	99US-0160770.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160814.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160815.
PR	26-JUL-1999;	99US-0145276.	PR	22-OCT-1999;	99US-0160980.
PR	27-JUL-1999;	99US-0145513.	PR	25-OCT-1999;	99US-0160981.
PR	27-JUL-1999;	99US-0145513.	PR	25-OCT-1999;	99US-0160981.
PR	28-JUL-1999;	99US-0145951.	PR	25-OCT-1999;	99US-0161404.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161405.
PR	02-AUG-1999;	99US-0146388.	PR	26-OCT-1999;	99US-0161406.
PR	03-AUG-1999;	99US-0147033.	PR	26-OCT-1999;	99US-0161360.
PR	04-AUG-1999;	99US-0147204.	PR	28-OCT-1999;	99US-0161361.
PR	05-AUG-1999;	99US-0147302.	PR	28-OCT-1999;	99US-0161921.
PR	05-AUG-1999;	99US-0147302.	PR	28-OCT-1999;	99US-0161922.
PR	06-AUG-1999;	99US-0147260.	PR	28-OCT-1999;	99US-0161993.
PR	06-AUG-1999;	99US-0147303.	PR	29-OCT-1999;	99US-0162142.
PR	09-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	10-AUG-1999;	99US-0147935.			
PR	11-AUG-1999;	99US-0148171.			
PR	12-AUG-1999;	99US-0148319.			
PR	13-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	16-AUG-1999;	99US-0148684.			
PR	17-AUG-1999;	99US-0149368.			
PR	18-AUG-1999;	99US-0149175.			
PR	20-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149723.			
PR	23-AUG-1999;	99US-0149929.			
PR	23-AUG-1999;	99US-0149902.			
PR	23-AUG-1999;	99US-0149930.			
PR	25-AUG-1999;	99US-0150566.			
PR	26-AUG-1999;	99US-0150884.			
PR	27-AUG-1999;	99US-0151065.			
PR	27-AUG-1999;	99US-0151066.			
PR	30-AUG-1999;	99US-0151080.			
PR	31-AUG-1999;	99US-0151303.			
PR	01-SEP-1999;	99US-0151438.			
PR	07-SEP-1999;	99US-0151930.			
PR	10-SEP-1999;	99US-0152363.			
PR	13-SEP-1999;	99US-0153070.			
PR	15-SEP-1999;	99US-0153738.			
PR	16-SEP-1999;	99US-0154018.			
PR	20-SEP-1999;	99US-0154039.			
PR		99US-0154779.			
<hr/>					
PR	22-SEP-1999;	99US-0155139.	PR	23-SEP-1999;	99US-0155486.
PR	23-SEP-1999;	99US-0155486.	PR	24-SEP-1999;	99US-0155659.
PR	26-SEP-1999;	99US-0156458.	PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.	PR	05-OCT-1999;	99US-0157753.
PR	07-OCT-1999;	99US-0157865.	PR	08-OCT-1999;	99US-0158029.
PR	12-OCT-1999;	99US-0158232.	PR	13-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.	PR	13-OCT-1999;	99US-0159294.
PR	14-OCT-1999;	99US-0159329.	PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.	PR	14-OCT-1999;	99US-0159337.
PR	18-OCT-1999;	99US-0159638.	PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.	PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.	PR	22-OCT-1999;	99US-0160814.
PR	22-OCT-1999;	99US-0160815.	PR	25-OCT-1999;	99US-0160980.
PR	25-OCT-1999;	99US-0160981.	PR	25-OCT-1999;	99US-0160981.
PR	25-OCT-1999;	99US-0161404.	PR	26-OCT-1999;	99US-0161405.
PR	26-OCT-1999;	99US-0161406.	PR	26-OCT-1999;	99US-0161360.
PR	28-OCT-1999;	99US-0161361.	PR	28-OCT-1999;	99US-0161921.
PR	28-OCT-1999;	99US-0161922.	PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.			
<hr/>					
Query Match					
Best Local Similarity 100.0%; Score 8; DB 21; Length 122;					
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	402 EDLSHNRH 409				
DB	76 EDLSHNRH 83				
<hr/>					
RESULT 12					
ID	AMG43679				
XX	AMG43679 standard; Protein; 122 AA.				
XX					
XX	AMG43679;				
XX					
XX	18-OCT-2000 (first entry)				
XX					
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 54622.				
XX					
XX	Protein identification; signal transduction pathway; metabolic pathway;				
XX	hybridisation assay; genetic mapping; gene expression control; promoter;				
XX	termination sequence.				
OS	Arabidopsis thaliana.				
PN	EP1033405-A2.				
XX					
XX	06-SEP-2000.				
PD					
XX	25-FEB-2000; 2000EP-0301439.				
PF					
XX	25-FEB-1999; 99US-0121825.				
PR	05-MAR-1999; 99US-0123180.				

Mon Dec 2 12:19:42 2002

us-09-868-760-7.oli.rag

PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 08-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-01300510.
PR 23-APR-1999; 99US-0130891.
PR 26-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134376.
PR 18-MAY-1999; 99US-0134941.
PR 19-MAY-1999; 99US-0135124.
PR 20-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136592.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137722.
PR 03-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138840.
PR 14-JUN-1999; 99US-0138453.
PR 16-JUN-1999; 99US-0139432.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 21-JUN-1999; 99US-0139899.
PR 22-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 23-JUN-1999; 99US-0140655.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 30-JUN-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142803.
PR 08-JUL-1999; 99US-0142920.
PR 09-JUL-1999; 99US-0142977.
PR 12-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144884.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145274.
PR 26-JUL-1999; 99US-0145716.
PR 27-JUL-1999; 99US-0145813.
PR 27-JUL-1999; 99US-0145819.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147028.
PR 04-AUG-1999; 99US-0147202.
PR 05-AUG-1999; 99US-0147280.
PR 05-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0148171.
PR 10-AUG-1999; 99US-0148319.
PR 11-AUG-1999; 99US-0148341.
PR 12-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149426.
PR 18-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 25-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.

```

PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159285.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 18-OCT-1999; 99US-0159388.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160742.
PR 21-OCT-1999; 99US-0160776.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 22-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 28-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 29-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match
Best Local Similarity 1.6%; Score 8; DB 21; Length 122;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 402 EDLSHNRH 409
DB 76 EDLSHNRH 83

RESULT 13
AAG08769
ID AAG08769 standard; Protein; 125 AA.
AC AAG08769;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SPQ ID NO: 6438.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121835.
XX 05-MAR-1999; 99US-0123160.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 25-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130445.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134376.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139494.
PR 18-JUN-1999; 99US-0139495.
PR 18-JUN-1999; 99US-0139496.
PR 18-JUN-1999; 99US-0139497.
PR 18-JUN-1999; 99US-0139498.
PR 18-JUN-1999; 99US-0139499.
PR 18-JUN-1999; 99US-0139450.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139869.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.

```

PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144332.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144684.
 PR 21-JUL-1999; 99US-0144614.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145112.
 PR 23-JUL-1999; 99US-0145113.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145219.
 PR 26-JUL-1999; 99US-0145213.
 PR 26-JUL-1999; 99US-0145214.
 PR 27-JUL-1999; 99US-0145318.
 PR 27-JUL-1999; 99US-0145319.
 PR 28-JUL-1999; 99US-0145317.
 PR 28-JUL-1999; 99US-0145386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 08-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150286.
 PR 26-AUG-1999; 99US-0150284.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 30-AUG-1999; 99US-0151293.
 PR 31-AUG-1999; 99US-0151298.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154038.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155113.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157573.
 PR 06-OCT-1999; 99US-0158029.
 PR 07-OCT-1999; 99US-0158265.
 PR 08-OCT-1999; 99US-0158322.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.

PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 21-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 22-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161359.
 PR 25-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 26-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161922.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 1.6%; Score 8: DB 21; Length 125;
 Best Local Similarity 100.0%; Pred. No 9.4; Mismatches 0;
 Matches 8; Conservative 0; Indels 0; Gaps 0;

OY 402 EDLSHRH 409
 DB 79 EDLSHRH 86

RESULT 14
 AAC43678 standard; Protein: 125 AA.
 ID AAC43678;
 AC AAC43678;
 XX
 XX 18-OCT-2000 (first entry)
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 54621.
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 54621.
 KW Protein identification: signal transduction pathway; metabolic pathway;
 KW hydrolase; assay: genetic mapping; gene expression control; promoter;
 KM termination sequence.
 OS Arabidopsis thaliana.
 OS Arabidopsis thaliana.
 PN EPI033405-A2.
 XX
 XX 06-SEP-2000.
 PD
 PD 06-SEP-2000.
 XX
 XX 25-FEB-2000; 2000EP-0301439.
 PF
 PF 25-FEB-1999; 99US-0121825.
 PR 03-MAR-1999; 99US-0123160.
 PR 03-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 23-MAR-1999; 99US-0126264.
 PR 23-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.

PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	07-MAY-1999;	99US-0132487.
PR	11-MAY-1999;	99US-0132863.
PR	14-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134321.
PR	18-MAY-1999;	99US-0134370.
PR	19-MAY-1999;	99US-0134768.
PR	20-MAY-1999;	99US-0134941.
PR	21-MAY-1999;	99US-0135124.
PR	24-MAY-1999;	99US-0135353.
PR	25-MAY-1999;	99US-0135639.
PR	27-MAY-1999;	99US-0136021.
PR	28-MAY-1999;	99US-0136392.
PR	01-JUN-1999;	99US-0136782.
PR	03-JUN-1999;	99US-0137222.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138867.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	21-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139889.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140685.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140981.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142305.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142940.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144375.
PR	19-JUL-1999;	99US-0144376.
PR	19-JUL-1999;	99US-0144382.
PR	19-JUL-1999;	99US-0144383.
PR	19-JUL-1999;	99US-0144384.
PR	20-JUL-1999;	99US-0144385.
PR	20-JUL-1999;	99US-0144386.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	03-AUG-1999;	99US-0146389.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	10-AUG-1999;	99US-0147935.
PR	11-AUG-1999;	99US-0148171.
PR	12-AUG-1999;	99US-0148319.
PR	13-AUG-1999;	99US-0148321.
PR	13-AUG-1999;	99US-0148365.
PR	16-AUG-1999;	99US-0148684.
PR	17-AUG-1999;	99US-0149368.
PR	18-AUG-1999;	99US-0149175.
PR	20-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158380.
PR	12-OCT-1999;	99US-0159283.
PR	13-OCT-1999;	99US-0159284.
PR	14-OCT-1999;	99US-0159285.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	18-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.

```
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.
```

```
Query Match 1.68; Score 8; DB 21; Length 125;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 402 EDLSHHRH 409
DB 79 EDLSHHRH 86
```

```
RESULT 15
AAU45288
ID AAU45288 standard; Protein: 259 AA.
```

```
XX AAU45288;
```

```
DT 27-FEB-2002 (first entry)
```

```
DE Propionibacterium acnes immunogenic protein #6184.
```

```
XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
```

```
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
```

```
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
```

```
XX dermatological; osteopathic; neuroprotectant.
```

```
OS Propionibacterium acnes.
```

```
XX MO200181581-A2.
```

```
XX 01-NOV-2001.
```

```
XX 20-APR-2001; 2001MO-US12865.
```

```
XX 21-APR-2000; 2000US-199047P.
```

```
XX 02-JUN-2000; 2000US-208841P.
```

```
XX 07-JUL-2000; 2000US-216747P.
```

```
XX (CORI-) CORIXA CORP.
```

```
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
```

```
XX L'maisonneuve J, Zhang Y, Jen S, Carter D;
```

```
XX WPI; 2001-616774/71.
```

```
XX N-PSDB; AAS59525.
```

```
XX Propionibacterium acnes polypeptides and nucleic acids useful for
```

```
XX treating acne vulgaris -
```

```
XX Example 1: SEQ ID No 6483; 1069pp; English.
```

```
XX Sequences AAU9105-AAU68017 represent Propionibacterium acnes immunogenic
XX polypeptides. The proteins and their associated DNA sequences are used in
XX the treatment, prevention and diagnosis of medical conditions caused by
XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
```

```
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The invention
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
```

```
XX Sequence 259 AA;
```

```
Query Match 1.68; Score 8; DB 22; Length 259;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 121 DLGDLVDD 128
DB 105 DLGDLVDD 112
```

```
Search completed: November 29, 2002, 12:44:41
Job time : 37 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2002, 07:50:58 ; Search time 2055 Seconds

(without alignments)
11750.594 Million cell updates/sec

Title: US-09-868-760-6

Perfect score: 1491

Sequence: 1 gaggggggagcagctgaaga.....gctgtgtataggaagagca 1491

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapept 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthm:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gq_est1:*
10: gq_est2:*
11: gq_hic:*
12: gq_est3:*
13: gq_est4:*
14: gq_est5:*
15: em_estfun:*
16: em_estom:*
17: gq_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65.6	4.4	908	17	A2548467 ENTKE30TR
2	64.2	4.3	877	17	A2531291 ENTBO34TR
3	64.2	4.3	912	17	A2551092 ENTET22TR
4	63.4	4.3	468	12	BG553005 ddb82e09.
5	62.2	4.2	795	17	A2528485 ENTOM64TR
6	62.2	4.2	823	17	A2676218 ENTKE36TR

7	62.2	4.2	843	17	BH139532
8	58.8	3.9	740	10	AV679268
9	58.4	3.9	976	17	BH149983
10	57.2	3.8	843	17	A2531618
11	57.2	3.8	887	17	B0731479
12	56.6	3.8	825	13	B374097
13	56.6	3.8	644	14	B0834753
14	56.4	3.8	661	14	B0834718
15	56.4	3.8	677	13	BH161314
16	56	3.8	782	17	A2669474
17	55.6	3.7	840	10	AV674403
18	55.4	3.7	607	13	B3348812
19	55.2	3.7	900	17	A2549980
20	55.5	3.7	906	17	BH153606
21	54.8	3.7	816	17	A2535744
22	54.8	3.7	891	17	A2683582
23	54.6	3.7	825	13	B374097
24	54	3.6	751	13	B374223
25	53.2	3.6	890	17	BH146886
26	52.8	3.5	493	13	BH170559
27	52.8	3.5	880	17	A2529191
28	52.6	3.5	931	17	BH160272
29	52.4	3.5	890	17	A2530768
30	52.2	3.5	548	13	BH159818
31	52	3.5	849	17	A2546009
32	51.8	3.5	445	17	BH167469
33	51.8	3.5	723	13	B339844
34	51.2	3.4	598	13	BH170666
35	51.2	3.4	605	13	BH163520
36	51.2	3.4	645	13	BH165350
37	51.2	3.4	647	13	BH163120
38	51.2	3.4	671	13	BH160252
39	51.2	3.4	694	14	BH168242
40	51.2	3.4	717	13	BH160500
41	51.2	3.4	747	13	BH162732
42	51.2	3.4	756	13	BH162492
43	51.2	3.4	764	13	B3398927
44	51.2	3.4	794	13	BH159906
45	50.8	3.4	323	13	BH171402

ALIGNMENTS

RESULT 1
A2548467
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

A2548467 908 bp DNA linear GSS 14-NOV-2000
ENTKE30TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
A2548467
A2548467.1 GI:11172102
GSS.
Entamoeba histolytica.
Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
1 (bases 1 to 908)
Determination of clone end sequences from Entamoeba histolytica
Loftus, B., Van Aken, S. and Fraser, C.
HM1:IMSS sheared DNA library
Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b1loftus@igf.org

Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library
Seq primer: M3-Reverse
Class: shotgun
High quality sequence start: 17
High quality sequence stop: 828.

FEATURES
SOURCE

```

FEATURES
SOURCE
1. 908
location/Qualifiers
/organism="Entamoeba histolytica"
/strain="HML:IMS5"
/db_xref="taxon:5759"
/clone.lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pHOSt; Site:1: Bat I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G. and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Battelli, Oxford University Press, 1999)."
BASE COUNT
434 a
16 c 289 g 165 t

```

Query Match	4.48;	Score 65.6;	DB 17;	Length 908;
Best Local Similarity	43.88;	Pred. No. 1e-06;		
Matches 336;	Conservative	0;	Mismatches 429;	Indels 3;
				Gaps 1;

QY 734 TAGAGGATTTAATGTANGTGTGCACCAAAAGATCATCTCCATGACG 78

Db 740 ATGATGAGATCATGACGAACACTGAAGATGATGAAGATCATGAAGAAG 78

RESULT 2	AZ531291	877 bp	DNA	linear	GSS 03-NOV-2000
LOCUS	AZ531291/c				
DEFINITION	ENT03044TR <i>Entamoeba histolytica</i> sheared DNA			<i>Entamoeba histolytica</i>	
ACCESSION	AZ531291				
VERSION	AZ531291.1	GI:1108538			
KEYWORDS	GSS				
SOURCE	<i>Entamoeba histolytica</i> .				
ORGANISM	<i>Entamoeba histolytica</i> .				
REFERENCE	1 (bases 1 to 877)				
AUTHORS	Loftus, B., Van Aken, S., and Fraser, C.				
TITLE	Determination of clone end sequences from <i>Entamoeba histolytica</i>				
JOURNAL	HMLIMS sheared DNA library				
COMMENT	Unpublished (2000)				
	Contact: Brendan J Loftus				
	<i>Entamoeba histolytica</i> Genomics				

BASE COUNT	156 a	266 c	29 g	424 t
ORIGIN				
Query Match		4.38;	Score 64.2;	DB 17;
Best Local Similarity		43.7%;	Pred. No. 2.4e-06;	Length 877;
Matches 276; Conservative		1;	Mismatches 354;	Indels 0; Gaps

QY	1	GAGGGACACAGTGTAAACGATGGGCGAAGCAAAAGTGTGCCACCCATGACACACAT	10
Db	776	GATATCATGATGAGGAAGGAAGGAAGATGATGAAGATGTGTGTGTAAGAAATGATGATAT	717
QY	61	GATACCATGTACGACCATGATGATGATGATGAAGAAACAAATCCATCTTGTGCCAGTGTAAATG	120
QY	716	GAAAGAAATATATATGAAAGAAAGATGATGATGATGATGAAGAAAGATGTGTAAGAAAGATGTGAA	657
Db			
QY	121	GACCAAAACCCCTCATGTGCTACGACGCTTCACCAACCAATGTCCATGTGCACACATAGAGTTG	180
QY			
Db	656	GAAAGAAAGAAATATATTAAGATGAAGATGATGATGAGAGACGATGAAGACGACGAATATGTGAATTA	597

[illegible]

method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is detailed in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).

[illegible]

ORGANISM	Entamoeba histolytica
REFERENCE	Eukaryota: Entamoebidae: Entamoeba.
AUTHORS	1 (bases 1 to 795)
TITLE	Loftus, B., Van Aken, S. and Fraser, C. Determination of clone end sequences from Entamoeba histolytica HML1:IMSS sheared DNA library Unpublished (2000)
JOURNAL	Contact: Brendan J Loftus Department of Eukaryotic Genomics The Institute for Genomic Research 3712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0208 Fax: 301 838 3543 Email: bjo@loftuseligf.org
COMMENT	Clones are derived from the Entamoeba histolytica HML1:IMSS sheared DNA library Seq primer: M13-Forward Class: shotgun High quality sequence start: 23 High quality sequence stop: 790. Location/Qualifiers 1..795
FEATURES	source 1..795 /organism="Entamoeba histolytica" /strain="HML1:IMSS" /db_xref="taxon:5759" /clone_lib="Entamoeba histolytica Sheared DNA" /note="Vector: pHOSt; Site 1: Bat I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + l method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaundin and B. Barell. Oxford University Press, 1999)."
BASE COUNT	319 a 72 c 175 g 229 t
ORIGIN	
Query Match	4.2% Score 62.2; DB 17; Length 795;
Best Local Similarity	47.5% Pred. No. 7.9e-06;
Matches 184; Conservative 0; Mismatches 203; Indels 0; Gaps 0;	
32	ACAGTGCACCACTGACGACACCAACGATGATCACCATGACGACCATGATGATGATG 91
317	ATGATGATATATGATATTTAGTATGATGACGATGATGATGATGATGATGATGATG 376
92	AAACAAATGCACATATGCCACAGTGTGAAATGAGAACCAACCCCTCATATGGCTACGAGCTTC 151
377	ATGATGATGATGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATG 436
152	ACCAACCATGTCATGACGACATAGAGTTGTACAGAAAGGCTCATGAGCTGTTTATCTAG 211
437	ATGACATGATGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 496
212	AACCTGATCTTTCGCGATTCACACACAAAGTGAAGACCACTGACGACCAACCATCATGCACTTC 271
497	ACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 556
272	ATCTGCACATGCTGTGTCATGTCAGAGAGTGTGTCATGATTCATTTGGGCGAACCTGATACAG 331
557	ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 616
332	CTGACCCAGAAAAAATCTGACCTCGGTGACCTCGGTGACACCTGTTGAGTGAAGGATGAGG 391
617	ATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 676
392	GCCTGGTTAATGAACTTCATCTATG 418
677	ATGATGATGACGATGATGATGATGATG 703

[illegible][illegible]

Mon Dec 2 12:19:39 2002

us-09-868-760-6.rst

Page 6

D _b	284	ATATGATGTAGTGTACCGATGATGTCAGTGGATGTGACGATGATGCATGATGATG	343
Q _y	152	ACCACCATCTCCATGGCAGCAATAGAAGTTGTCCAGAAAGSGTCATGGAGCTGTATTCTAG	211
D _b	344	ATGACGATGTACGATGTGATGTGATGATGATGATGATGATGATGATGATGATGATGATG	403
Q _y	212	AACCTCATCTTGTCGATTTCAACACAGATGAAGCACATGACGACCATCATGACTTC	271
D _b	404	ACGATGATGTGTACGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATG	463
Q _y	272	ATCTGCACATGCTTGCTGTGCATCTGCACAGTGTGTATTCTATTGGGACACTGTACATG	331
D _b	464	ATGATGACGATGTGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	523
Q _y	332	CTCACCCAGAAAACAACCTGCTGACCTCGSTGACCTTCGCTGCTGTGACAGTAAGGG	391
D _b	524	ATGACGATGTGTGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	583
Q _y	392	GCGTGTATTATGAAGTTCATCAATTG	418
D _b	584	ATGATGATGACGATGAATATGATATG	610

RESULT	8
AV679268/c	
LOCUS	740 bp mRNA linear EST 05-OCT-2000
DEFINITION	AV679268 Nori Satoh unpublished cDNA library Clona Intestinalis
ACCESSION	AV679268
VERSION	cdna clone rc0tbtsh19.3', mRNA sequence.
KEYWORDS	AV679268
SOURCE	AV679268.1 GI:10117267
ORGANISM	EST.
	Clona intestinalis.
	Clona intestinalis.
	Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
	Phlebobranchia; Clonidae; Clona.
REFERENCE	1 (bases 1 to 740)
AUTHORS	Satoh,N., Satou,Y., Kohara,Y. and Shin-I,T.
TITLE	Expressed genes in clona intestinalis
JOURNAL	Unpublished (2000)
COMMENT	Contact: Nori Satoh

FEATURES
 KYOTO UNIVERSITY
 Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
 Tel: 81-75-753-4081
 Fax: 81-75-705-113
 Email: satchos@cidian.zool.kyoto-u.ac.jp
SOURCE
 Location/Qualifiers
 1..740

FEATURES	source	Location/Qualifiers
	1. .740	/organism="Clona intestinalis"
		/db_xref="taxon:7719"
		/clone="rc1b15h19"
		/clone_1fb="Mori Satoh unpublished cdna library"
		/tissue_type="whole animal"
		/dev_stage="tailbud"
		/note="Vector: pBluescript SK"
BASE COUNT	150 a	216 c 68 g 292 t 14 others
ORIGIN		

	Query Match	3.9%	Score 58.8	DB 10	Length 740
	Best Local Similarity	45.7%	Pred. No. 6.5e-05		
	Matches 195	Conservative 0	Mismatches 222	Indels 0	Gaps 0
QY	12	GTTAGAGATGGCGCAGCAAGATGATACCAACATGACGACACCAACGATGATACCATGA	71		
Db	550	GNGAGATGATACCTAGATGATATACCAAGATGATGNNAGATGTGTCGCAAGTGA	491		
QY	72	CGACCATGATGATGATGATGAACCAATGCACTATGCTCCCATGCTGAATGNAACCAACC	131		
Db	490	CGAAATGATATACCAACATGATGATGACCAAGTATGATGCGCAAGTATGA	431		
QY	132	TCATATGGCTAGCAGCCTTACCAACCATGTCATGCTGACGACATAGAGTTGTACAGAAAGG	191		

[illegible]

RESULT	9
BHLI49983	
LOCUS	BHLI49983
DEFINITION	BHLI49983 976 bp DNA linear GSS-27-AUG-2001 ENTQ09J3FF Entamoeba histolytica Sheared DNA Entamoeba histolytica genomicC, DNA sequence.
ACCESSION	BHLI49983
VERSION	BHLI49983.1 GI:15311165
KEYWORDS	GSS.
SOURCE	Entamoeba histolytica. Entamoeba histolytica Eukaryota; Eutrombozoa; Eukaryota; Eukaryota; Eukaryota; 1 (bases 1 to 976)
ORGANISM	Loftus,B., Wang,Z., Van Aken,S. and Fraser,C. Determination of clone end sequences from Entamoeba histolytica IM1:IMS sheared DNA library (2001) Unpublished (2001) Contact: Brendan J Loftus Department of Eukaryotic Genomics
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel.: 301 838 0208
Fax: 301 838 3543
Email: bjloctus@igf.org
Clones are derived from the *Entamoeba histolytica* HMI-TMS sheared
DNA library
Seq Primer: M3-Forward
Class: Shotgun
High quality sequence start: 38
High quality sequence stop: 740.
Location/Qualifiers

```

/organism="Entamoeba histolytica"
/strain="HM:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: PHOS1; Site: 1; Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Battell. Oxford University Press, 1999)."
```


RESULT	11
LOCUS	B0731479
DEFINITION	BG0731479 887 bp mRNA linear EST 16-JUL-2002 AGENCOUNT_8216778 NICHD XCC Embd4 Xenopus laevis cDNA clone IMAGE:4930330 5'', mRNA sequence.
ACCESSION	B0731479
VERSION	B0731479.1 GI:21870376
KEYWORDS	EST.
SOURCE	African clawed frog. Xenopus laevis
ORGANISM	Xenopus laevis Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae; Xenopodinae; Xenopus. 1 (bases 1 to 887)
REFERENCE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
AUTHORS	Tumor Gene Index
TITLE	Unpublished (1997)
JOURNAL	Contact: Robert Stransberg, Ph.D. Email: robert.stransb@nih.gov
COMMENT	

Tissue Procurement: Dr. Igor Dawid
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: NCI-CCSP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bcrp/image/image.html
Plate: LLAM10852 row: 1 column: 11
High quality sequence stop: 603.

FEATURES	SOURCE	Location/Qualifiers
		1. 887
		/organism="Xenopus laevis"
		/db_xref="taxon:8355"
		/clone="IMAGE:4930330"
		/clone_lib="NICHD XGC Emb4"
		/dev_stage="embryo, stage 31-32"
		/lab_host="PH10B (phage-resistant)"
		/note="Organ: Whole embryo; Vector: PCMV-SPOB6; Site: 1;
		Not1; Site: 2; Salt: Cloned unidirectionally. Primer: oligo
		dr. Average insert size 2.1 kb. Constructed by Life
		Technologies. Note: This is a Xenopus Gene Collection (XGC
) Library."
BASE COUNT	268 a	172 c 242 g 205 t
ORIGIN		

	Query Match	Best Local Similarity	3.8%; 46.9%;	Score 57.2; DB 14	Length 887;
	Matches 179;	Conservative 0;	Mismatches 203;	Indels 0;	Gaps 0
QY	32	AAGTATGCCACCACTGACGACACCAACGATGATACCACTGACGACGACCATGATGATGATG			
DB	350	ACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG			
QY	92	AAACATATGCACATATGCCCCATGTGAAATGGAACCAACCCATATATGGCTTGTGACGACCTTC			
DB	410	ATGATGACGACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG			
QY	152	ACCAACATATGCATATCCAGATGACATGATGATGTCTACACAGATGATGACGATGATGATGATG			
DB	470	ACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG			
QY	212	AACTTCATCTTTTGTGGATTCATACACAATGTAAGAACCATGATGACGACCATATATGAGACTTC			
DB	530	ACGATG			
QY	272	ATCTGCACATGCTTGTGTGATCATCTACAGCAGGTTTGTGATTTCTATTGCGCACTGTACCAATG			
DB	590	TACAAAAGATGACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG			
QY	332	CTCACCCAGAAAACATGCTGACCCCTGTGTACCTCGGTGATGACCTGGTTGACGATGATGAGG			
DB	650	ATGACGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG			

Qy 392 GCGTGGTTAATGAAGTTCATCA 413
| | | | | | |
Db 710 ATGACCAGAAAGATGATGATGA 731

RESULT 12				
Bj374097				
LOCUS				
DEFINITION				
DESCRIPTION				
Bj374097	825 bp	linear	EST 08-MAR-2002	
Bj374097	Dictyostelium discoideum	CDNA library, CF	Dictyostelium	
discoidium	cdna clone ddcdb15 3',	mrna sequence.		
Bj374097				

ACCESSION	B0374069
VERSION	B0374069.1
KEYWORDS	GI:19283480
SOURCE	EST.
ORGANISM	Dicystostellum discoideum.
REFERENCE	Dicystostellum discoideum.
AUTHORS	Eukaryota; Mycetozoa; Dictyostelida; Dictyostellium.
TITLE	1 (bases 1 to 825)
COMMENT	Crushihara, H., Tanaka, Y., Kohara, Y. and Shira-i T. Full length cDNA of Dicystostellum discoideum at the culmination stage

JOURNAL: Unpublished (2002)
 COMMENT: Contact: Tadasu Shin-1
 Center for Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6636
 Fax: 81-559-81-6635
 Email: tshin1@genetics.nig.ac.jp

```

FEATURES
  source
    location/Qualifiers
      1. .825
        /organism="Dictyostellium discoideum"
        /strain="AX4"
        /db_xref="taxon:44689"
        /clone="ddc6b15"
        /clone_1b="Dictyostellium discoideum cDNA library, CF"
        /sex="mat A"
        /dev_stage="Culmination stage"
  BASE COUNT
    190 a 113 c 133 g 229 t 160 others
  ORIGIN

```

Query Match	3.8%;	Score 56.8;	DB 13;	Length 825;
Best Local Similarity	36.1%;	Pred. No. 0.00026;		
Matches 259; Conservative	0;	Mismatches 451;	Indels 7;	Gaps 2;

QY 203 TTATATCAGAACTTCATCTTGGCGATTTCAACACAAGAGGAGACCATGAGACACCATG 262

Db 96 TTGACCTAGAACTGCTTGAGCTAGAACTCTTGAGCTTGAACCTTGAGCTGCTTGAGC 155

QY 263 ATGACCTTCATCTGCACATGCTTGGTGACATCTGACGACGGTTTGATTTATTGGCGAAC 322

Db 156 TTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 215

QY 323 TGTCATCATGCTCAACCCAGAAAACATGCTGACCTCGTGACCTCGTGACCTGGTTGAGC 382

Db 216 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 275

QY 383 ATGATAGGGGCGCTGTTAATGAAGTCATCATATTGCTTGGTTGGACCTTGATGGT--AC 440

Db 276 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 335

QY 441 AGCACCAACACGAGACCTCTCATTTGGACATCTCAATGACTATTTTACAAGGAGTGCAC 500

Db 336 GCGCNCNCTGANNAGNAGCNCNNNNNCNCTGANGAGNAGCNCNNCNCNCTCNCNTNANN 395

QY 501 CGATGCTGATACCCACGCGAGTGAATTCGCCCTGGTGTGTTATTGTCATGAGAAAAGCTG 560

Db 396 ANCGACNNCNCNNNAGNANCNTNANCNCCTGCTCNCNNCNCNAGNANNNTNANCNNCTG 455

QY 561 CCCGAAAGACGACGCTGCTCTTACATACGAGGTAGAGAGATYAAAACGCTGAGCTTTATG 620

Db 456 NTNATCANCNC-----NNCNCNCTAGNANCNTNANCACACTCNCNNNNCNCNAGNANCN 510

QY 621 CCAATTTGACGTAAGATCTAATACACACCAACCAACCAAGGCTTCTATCATCATATGTCACG 680

Db 511 TCNTCNTNANCAACNNCNCACNNCTCCTCNCNCNTNANNANCAACNNCNCNT 570

Qy 681 AACCATGCATTTGCAACACAGTGGTTGTTATGGACCTTGAAGTGTCTCCATCTTAGAGGG 740

Db 571 NCNTCNTNTNATGATGAACAACTCNCCTCTTNNATNATGATGANNATCATGATGATGATNA 630

Qy 741 ATTTAATGTAAGTGATGACCAACAAAGTCATCTCATGACGTACAGATCTACGCCAAGG 800

Db 631 NCNTCNTTTATGATGATGATGATNATNANCCTCNCNATNANCNTCNCNTNTNTGATNA 690

Qy 801 TGACCTGCACAGTGGATGATPAACCTGGGTGTAATATGATCCTCATGAAGATTACCA 860

Db 691 TNAATCTCNCNTCCTTMAATNATNATCATCATCANCNTCNTMTNATGATGATGATCA 750

Qy 861 CAGTGATGGGTGATCTGAGAGATATTCACAGATGAGACATGGGGTTCATGCA 917

Db 751 TCATCNCNTNATGATGATCATCANCCTCNCCTNATGATGATGATGATGATGATGA 807

RESULT 13	
BO834753/C	
LOCUS	644 bp mRNA linear EST 08-AUG-2002
DEFINITION	Po.ad.03M07.TEXF1 Psoroptes ovis mixed Psoroptes ovis CDNA clone
ACCESSION	Po.ad.03M07 5', mRNA sequence.
VERSION	BO834753
KEYWORDS	BO834753.1 GI:22139067
SOURCE	EST.
	sheep scab mite.

REFERENCE	1 (bases 1 to 644)
AUTHORS	Kenyon, F., Welsh, M., Parkinson, J., Whilton, C., Blaxter, M. and Knox, D. P.
TITLE	Expressed sequence tag survey of gene expression in the scab mite <i>Psoroptes ovis</i> - allergens, proteinases and free radical scavengers
JOURNAL	Unpublished (2002)
COMMENT	Contact: Blaxter ML

Institute of Cell, Animal and Population Biology
University of Edinburgh
Aspathorn Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JF, UK.
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
The library was prepared for Dr David Knox, Moredun Research
Institute, UK Sequencing was performed by Mark Welsh, ICAPB,
Edinburgh
PCR Primers
FORWARD: M13R
BACKWARD: T7PL
Plate: 03 row: A column: 07
High quality sequence stop: 536.

```

FEATURES
    1. 644
    Location/Qualifiers
        /organism="Psoroptes ovis"
        /db_xref="taxon:83912"
        /clone_1b="Po.ad.03A07"
        /clone_1lb="Psoroptes ovis mixed"
        /sex="Mixed"
        /dev_stage="Mixed (predominantly adult)"
        /note="Vector: lambda Triplex2, Site_1: SfilA, Site_2:
        SfilB, Psoroptes ovis is the infective agent causing sheep
        scab, a serious dermatitis in sheep. The cDNA was
        obtained from miles harvested from the skin of previously
        unexposed sheep harbouring a moderate challenge
        infection. The cDNA was cloned using the Clontech lambda
        Triplex2 Smart cDNA cloning system."
BASE COUNT
    205 a      133 c      123 g      183 t
ORIGIN

```

Query Match	3.88;	Score 56.6;	DB 14;	Length 644;
Best Local Similarity	47.6%;	Pred. No. 0.00024;		
Matches 167; Conservative	0;	Mismatches 184;	Indels 0;	Gaps 0;

QY	607	ACTGAGCATTAATGGCCAAATTGTGACGTATGAATCTAATAACACCACAACCAAGGCCTTTCAT	666
Db	496	ACCAAAGAAGCATGTTATGTGACATGACGACGATGATGATGCTACTCCTCAATATTAATGAATGGC	437
QY	667	CATCATGTGCCAGGAACCATCGATTCTTCAACAAAGTTGGTTATGTGTGACCTTGAAAGTGCC	726
Db	436	CACCATGTGACCATGTGATGATGATATACATAAAGTTGTTGACCATGATGTGCTCAATGGTGTATGTC	377
QY	727	TACCATTTTAGAGGATTTTAATGATTAAGTATSATCACCAAAAGATCATCTCCATGACGTACAG	786
Db	376	CAGGTGATGACGCCACGCGTCCCCCATTAATGTCCACCATGATGTCCAGGAATGAATGATAT	317
QY	787	ATCTACGCCAACGGGTGACCTGACAGTAGATGATTAACCTGGGTGCTAAATATGATGCTT	846
Db	316	GAGGACCTCAACATGTGCCACCATGATGATGTCCATGATGATATACCGGAGGTGCTCGGCCAATGATGTC	257
QY	847	CATGAAGATTATCCACAGATGATGGTGGATCTAGAGATATTTCCAGATGATGACCAATGGC	906
Db	256	CATGGGTGATGCCGCCAGTAGCCGACATATCTCTATGTGTGGTAGGCCACGAGGACCAAT	197
QY	907	GTTGTCAATGAACCCACAGATATTTCTGGATCAATATCTTCGGTGTATGAC	957
Db	196	GATGACCCACGAAAATGACACCATGGTCAATCGAATATACACGCCATGATGATTAAC	146

RESULT ID	LOCUS	DEFINITION	ACCSSION VERSION	KEYWORDS	SOURCE ORGANISM
B0834718/c	B0834718	661 bp mRNA	1linear	EST 08-AUG-2002	
	PO-4d.02x05	TEXF1 Psoroptes ovis mixed Psoroptes ovis CDNA clone			
	PO-4d.02x05	5', mRNA sequence.			
	B0834718.1	GI:22139032			
	EST				
	sheep scab mite.				
	Psoroptes ovis				
	Euryarchaeota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;				

REFERENCE	Acariformes; Sarcopitiformes; Astigmata; Sarcopitoidea; Psoroptidae; Psoroptes.
AUTHORS	1 (bases 1 to 661) Kenyon, F., Welsh, M., Parkinson, J., Whitton, C., Blaxter, M. and Knox, D. P.
TITLE	Expressed sequence tag survey of gene expression in the scab mite Psoroptes ovis allergens, proteinases and free radical scavengers unpublished (2002)
JOURNAL	CONTACT: Blaxter, M.L. Institute of Cell, Animal and Population Biology University of Edinburgh Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JF, UK. Tel: +44 131 650 6760 Fax: +44 131 670 5450

```

Email: mark.blaxter@ed.ac.uk
The library was prepared for Dr David Knox, Moredun Research
Institute, UK Sequencing was performed by Mark Welsh, ICAPB,
Edinburgh
PCR primers
FORWARD: M13R
BACKWARD: T7PL
Plate: 02 row: F column: 05
High quality sequence stop: 533.
Location/Qualifiers
1..661
/organism="Psoroptes ovis"
/db_xref="taxon:83912"
/clone_lib="Po.ad.02F05"
/clone_lib="Psoroptes ovis mixed"
/sex="Mixed"
/dev_stage="Mixed (predominantly adult)"
/note="Vector: Lambda TriPLEX2; Site:1: SfilA; Site:2:

```

slf1b; psoroptes ovis is the infective agent causing sheep scab, a serious dermatitis in sheep. The cDNA was obtained from mites harvested from the skin of previously unexposed sheep harbouring a moderate challenge infection. The cDNA was cloned using the Clontech Lambda Triplex2 Smart cDNA cloning system."

BASE COUNT 211 a 136 c 130 g 184 t
ORIGIN
Query Match 3.8%; Score 56.6; DB 14; Length 661;
Best Local Similarity 47.6%; Pred. No. 0.00024;
Matches 167; Conservative 0; Mismatches 184; Indels 0; Gaps 0;
OY 607 ACTGAGCATTTGCCCTGTGACGTAAAGTATATACACACCAAGGCTCTTCAT 666
DB 505 ACCAAGAGCATGTTTGTGATGATGATGATGATGATGATGATGATGATGATG 446
OY 667 CATCATGTCCACGACCATCATGATTTCAACAGTGTGTATGTCGTCCTGAAGTCC 726
DB 445 CACCATGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 386
OY 727 TACCATTTAGAGGATTTAATGTAAGTATGACACCAAGATCATCTCCATGACGTAC 786
DB 385 CAAAGTGTATGACCCGACAGTCCCATTAATGTCACCATGATGTCACGATGATGAT 326
OY 787 ACTACGCCAACGGTGAACCTGACAGTGTATGATTAACCTGCTTAATGATGATCT 846
DB 325 GAGGACATCAAGTGTCCACCATGATGATGATGATGATGATGATGATGATGATG 266
OY 847 CATGAAGATTACACAGTATGTTGGTATCTAGAGATATTCACGATGATGATGATG 906
DB 265 CATGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 206
OY 907 GTGTCAATGAAGAAGCCAGATATTCCTGGATCATATCTTCTGGTATGATGAC 957
DB 205 GATGACACGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 155
RESULT 15 677 bp mRNA linear EST 04-DEC-2001
BM161314 EST63837 PyBS Plasmodium yoelli yoelli cDNA clone PYCKD34 5' end,
LOCUS mRNA sequence.
DEFINITION BM161314 GI:17306995
ACCESSION
VERSION EST.
KEYWORDS Plasmodium yoelli yoelli.
SOURCE Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ORGANISM
REFERENCE 1 (bases 1 to 677)
AUTHORS Carlom,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdya,A.B.,
Fraser,C.M. and Carucci,D.J.
Plasmodium yoelli EST Project at TIGR
TITLE Unpublished (2001)
JOURNAL Contact: Jane Carlom
COMMENT Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlont@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.
FEATURES
Location/Qualifiers
1..677
/organism="Plasmodium yoelli yoelli"
/strain="17XL"
/db_xref="taxon:73239"
/clone="PYCKD34"
/clone_lib="PyBS"
/dev_stage="Asexual blood stages"

/lab host="E. coli XL-1 Blue"
/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cBYJ mice infected with PY17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven adapters ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybrILAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybrILAP vector and plasmid DNA isolated."

BASE COUNT 288 a 50 c 186 g 153 t
ORIGIN
Query Match 3.8%; Score 56.4; DB 13; Length 677;
Best Local Similarity 48.4%; Pred. No. 0.00028;
Matches 156; Conservative 0; Mismatches 186; Indels 0; Gaps 0;
OY 598 GAAGATAAATGAGCATTTATGCCATTTGACGTAAATCTAATACACCAACCAAG 657
DB 91 GAAATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 150
OY 658 GCTCTTCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 717
DB 151 ACAAGGAAATGAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 210
OY 718 GAAGTCTCTTACCATTTAGAGGATTTAATGTAAGTATGATGATGATGATGATG 777
DB 211 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 270
OY 778 GAGGTACAGATCTAGCCACGAGTGTGACCTGACAGTGTGATGATTAACCTGGTCTAAA 837
DB 271 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 330
OY 838 TATGATCTTCATGAAGATTTACACAGTGTGGTATCTAGAGATTTTACCATGAT 897
DB 331 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 390
OY 898 GACCATGGCTGTCTCAATGA 919
DB 391 GACGATGATGATTTTGAAGATA 412

Search completed: November 29, 2002, 10:02:51
Job time : 2074 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2002, 07:43:13 ; Search time 2829 Seconds

(without alignments)
15338.382 Million cell updates/sec

Title: US-09-868-760-6

Perfect score: 1491

Sequence: 1 gaggggagcagtgtaacga.....gctgtatagagcagcga 1491

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402678 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank :
1: gb_ba :
2: gb_hlg :
3: gb_in :
4: gb_om :
5: gb_ov :
6: gb_pat :
7: gb_ph :
8: gb_pl :
9: gb_pr :
10: gb_ro :
11: gb_sts :
12: gb_sy :
13: gb_un :
14: gb_vl :
15: em_ba :
16: em_fun :
17: em_hum :
18: em_in :
19: em_mu :
20: em_om :
21: em_or :
22: em_ov :
23: em_pat :
24: em_ph :
25: em_pl :
26: em_ro :
27: em_sts :
28: em_un :
29: em_vl :
30: em_hlg_hum :
31: em_hlg_inv :
32: em_hlg_other :
33: em_hlg_mus :
34: em_hlg_pln :
35: em_hlg_rtd :
36: em_hlg_mam :
37: em_hlg_vtl :
38: em_sy :
39: em_higo_hum :
40: em_higo_mus :
41: em_higo_other :

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1484.2	99.5	1700	3	AF273766
2	88.4	5.9	115758	9	AC104634
3	74.8	5.0	186278	9	AC079176
4	69.6	4.7	272545	2	AC090533
5	66.4	4.5	143716	2	AL627248
6	65.2	4.4	222546	2	AC096302
7	64	4.3	272545	2	AC090533
8	63.6	4.3	163243	9	AL513304
9	59.6	4.0	211001	2	AC018695
10	58.4	3.9	61052	2	AC117074
11	58.4	3.9	61052	2	AC123513
12	58.4	3.9	163690	2	AC131509
13	57.2	3.8	51939	9	AC099341
14	56.8	3.8	176480	2	AC104805
15	56.6	3.8	388	8	PAB296740
16	56	3.8	1592	8	AF046872
17	55.6	3.7	160898	9	AC025540
18	55.5	3.7	186558	2	AC079031
19	54.8	3.7	163801	2	AC025224
20	54.6	3.7	26085	2	AC096530
21	54.4	3.6	172853	9	AC084361
22	54.2	3.6	210067	2	AC026358
23	54	3.6	205429	2	AC005506
24	53.8	3.6	262428	2	AC096326
25	53.6	3.6	171260	9	AC026743
26	53.6	3.6	49144	2	AC131380
27	53.4	3.6	164399	3	PFMAL3P6
28	52.8	3.5	59762	8	AB023032
29	52.6	3.5	91048	9	AC008619
30	52.6	3.5	157420	9	HS569D19
31	52.6	3.5	114771	2	AL845482
32	52.6	3.5	173287	2	AC104753
33	52.6	3.5	177070	2	AC018875
34	52.6	3.5	188121	2	AC012451
35	52.6	3.5	218815	2	AL807384
36	52.4	3.5	306	8	AF231791
37	52.4	3.5	550	11	G67741
38	52.4	3.5	60953	2	AC100470
39	52.4	3.5	142447	10	AL627204
40	52.4	3.5	254197	2	AF389853
41	52.2	3.5	115489	2	AC117072
42	52.2	3.5	186562	2	AL451079
43	52.2	3.5	200368	9	AC096541
44	52	3.5	83122	8	ATT20010
45	52	3.5	145524	2	AC016777

ALIGNMENTS

RESULT 1
LOCUS AF273766 1700 bp mRNA linear INV 20-MAR-2001
DEFINITION Perna canaliculus perna precursor, mRNA, complete cds.
ACCESSION AF273766
VERSION AF273766.1 GI:13383377
KEYWORDS
SOURCE
ORGANISM Perna canaliculus.
Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Mytiloidea;
Mytiloidea; Mytilidae; Perna.
REFERENCE
1 (bases 1 to 1700)
Scotti,P.D., Dearing,S.C., Greenwood,D.R. and Newcomb,R.D.
Perna: a novel, self-aggregating haemolymph protein from the New
Zealand green-lipped mussel, Perna canaliculus (Bivalvia:

Journal. Mytilidae) Comp. Biochem. Physiol. B, Biochem. Mol. Biol. 128 (4), 767-779 (2001)
Medline 21186417
Reference 11290459
Authors 2 (bases 1 to 1700)
Title Scotti, P.D., Dealing, S.C., Greenwood, D.R. and Newcomb, R.D. Direct Submission Submitted (31-MAY-2000) The Horticulture and Food Research Institute of New Zealand Ltd, 120 Mt. Albert Road, Auckland, New Zealand

FEATURES	Location/Qualifiers
source	1..1700
	/organism="Perna canaliculus"
	/db_xref="taxon:38949"
CDS	34..1587

/note="naemolymph protein; N-terminus determined by
 microsequencing of perlin; DEQGNNDGN and HPLC-purified
 CSH and tryptic digest fragments; Ass11NHVHG; VVNHVNH;
 GSHPEYVH; YHGHDHA; QGSHVEVSERVACVYIGRA"
 /codon_start=1
 /evidence=experimental
 /product="perlin precursor"
 /protein_id="AAK0952.1"
 /db_xref="GI:13393378"
 /translation="MKLLSLVLRALLDQVRAGEQSNQKRDYHNDHNDHND
 DDDDDDEVTYADQEMERPHASSSLAHNVHSGSTELSSQGNAGVUTLEHLYGFTSD
 HNDHNDHNDHLMELDMSAGDSITGELINAPREHAPRGDLVDDEGVNLEYNIA
 WLDLGTAEVTEALILGHSMITLQGSTIDYDIPKSRACCVLEGRKAEYALAHLE
 EDKDTAEHACDVRSNTHOPKALHNHNVHGTLEKQVGTIDHLEGVNEDQNDH
 LHLDDQVLYANGDITSGCDNLGAKITDHEHDSIGLADIPDDHGVNESHRSYMLN
 LFGDSSVIGRSIALIHQRDLHKSAKIALACSVTIGQSHSPETVIRACVYKREYESTGRI
 HNVSSITFEQTGGSTHMTADLKGVNSHNHNGVLEWQVLEWQSDNHSNGHSGRI
 HNDHDAHPRKRPDGLVDITDHSGLVHSTKFDHLNVEDLNARSLVIMQGHVESEK
 VACCVYIGRA"
 34. 93
 sig.peptide
 mat.peptide
 94. 1584
 /product="perlin"
 1650. 1655
 polyA-signal
 515 a 375 c 379 g 431 t
 BASE COUNT
 ORIGIN

	Query Match	99.5%	Score 1484.2	DB 3:	Length 1700;
	Best Local Similarity	99.7%	Fred. No. 0;		
	Matches 1486;	Conservative	1; Mismatches	4; Indels	0; Gaps
OY	I	GAYGGGAGCAGTGTATACGATGSGCGCAGAACAAAGATGACCCACCATGACGCACCCACGAT	60		
Db	94	GATGGCCMACACGTGTAAATGATGGCGAGAACAAMAGATGACCCACCATTGACGCACCCACGAT	153		
OY	61	GATCACCATCAGCACCATGATGATGATGATGAACAAATGACACTATGCCCGTGTAANA	120		
Db	154	GATCACCATCAGCACCATGATGATGATGATGATAAACAATGACACTATGCCCGTGTAANA	213		
OY	121	GAACCAAAACCTCATATGGCTGTGACGCCCTTCACCAACATGTGCATGGCAGCATAGAGTTG	180		
Db	214	GAACCAAAACCTCATATGGCTGTGACGCCCTTCACCAACATGTGCATGGCAGCATAGAGTTG	273		
OY	181	TCACAGAAAGGTCATGGAGCTGTTTATCTBAGACTTCATCTTGTCGATTCACACAAGT	240		
Db	274	TCACAGAGAGGTCATGAGAGCTGTTTATCTBAGACTTCATCTTGTCGATTCACACAAGT	333		
OY	241	GAAGACCATACGACCCACCATCATGACTTCATCTGCAACATGCTTGGTGCATGTCCAGA	300		
Db	334	GAAGACCATACGACCCACCATCATGACTTCATCTGCAACATGCTTGGTGCATGTCCAGA	393		
OY	301	GGTGTATATCTATNTGGCGAAGTCATGTCACCCAGAAAAACATGGCAGACCCGTGT	360		
Db	394	GCTGTATATCTATNTGGCGAAGTCATGTCACCCAGAAAAACATGGCAGACCCGTGT	453		
OY	361	GACCTCGATGACCTGTTGACGATGATAGGGCGTGTTAATGAAGTTCAATCATATGCT	420		
Db	454	GACCTCGATGACCTGTTGACGATGATAGGGCGTGTTAATGAAGTTCAATCATATGCT	513		

QY	421	TGGTTGGACATTGTA	TGTTGTAAGACACCAACCAACCGAAGCTC	CAATGAGACATCAATGACT	480
Db	514	TGGTTGGACATTGTA	TGTTGTAAGACACCAACCAACCGAAGCTC	CAATGAGACATCAATGACT	573
QY	481	ATTATTACAAGGAGCTCA	CCGATGCTGATATCCCAAGCAGATAGAAATCCGCTGTTGTGT		540
Db	574	ATTATTACAAGGAGCTCA	CCGATGCTGATATCCCAAGCAGATAGAAATCCGCTGTTGTGT		633
QY	541	ATTGTGCTATGAAAAAGCTG	CCCGCAGAACAAGCAGCTGCTTACATCAACGAGCTAGAGNAA	600	
Db	634	ATTGTGCTATGAAAAAGCTG	CCCGCAGAACAAGCAGCTGCTTACATCAACGAGCTAGAGNAA	693	
QY	601	GATAAAACTGAGACATTAT	ATGCCATTGTGACGTAAAGATCTAATACACCAACCAAGGCT	660	
Db	694	GATAAAACTGAGACATTAT	ATGCCATTGTGACGTAAATATATATACCAACCAAGGCT	753	
QY	661	CTTTCATCATCATATGTC	CAACGAGACCATCGATTTTCAAAACAAGTTGGTTATGTGTGACTTGA	720	
Db	754	CTTTCATCATCATATGTC	CAACGAGACCATCGATTTTCAAAACAAGTTGGTTATGTGTGACTTGA	813	
QY	721	GTGTCTTACATATTTAG	AGAGGATTATTAATGTAACTGATGACCAACAAGATCATCTTCCATGAC	780	
Db	814	GTGTCTTACATATTTAG	AGAGGATTATTAATGTAACTGATGACCAACAAGATCATCTTCCATGAC	873	
QY	781	GTACACATCTACGCCA	AGGAGACCTGACAGTATGATTAATCCTCGGTGCTAAATAT	840	
Db	874	GTACACATCTACGCCA	AGGAGACCTGACAGTATGATTAATCCTCGGTGCTAAATAT	933	
QY	841	GATCCCATCAAGAATTA	TACCAAGTGAAGTTGGGGATATGACGAATTTCAACAGATGATGAC	900	
Db	934	GATCCCATCAAGAATTA	TACCAAGTGAAGTTGGGGATATGACGAATTTCAACAGATGATGAC	993	
QY	901	CATGGCGCTGTCAATGA	AAAGCACAAGTATTCCTGATCATATCTTCGGTGAATGACAT	960	
Db	994	CATGGCGCTGTCAATGA	AAAGCACAAGTATTCCTGATCATATCTTCGGTGAATGACAT	1053	
QY	961	GTCTCTGGAGCTTCA	TATGCGCATCAACCAAAAGACACATCTTCAATAAAATGTGCAAAAT	1020	
Db	1054	GTCTCTGGAGCTTCA	TATGCGCATCAACCAAAAGACACATCTTCAATAAAATGTGCAAAAT	1113	
QY	1021	GCCCTGTTGTGTCA	TATGAGCCTGGACACAGCATCCAGAAATGTTCAACAGATCAAAATG	1080	
Db	1114	GCCCTGTTGTGTCA	TATGAGCCTGGAGCAGACAGCATCCAGAAATGTTCAACAGATCAAAATG	1177	
QY	1081	GTTGTCAAGACTAATACA	GAATCTACTGTGTTACATCACATGTCTCTGGTCTATTAACA	1144	
Db	1174	GTTGTCAAGACTAATACA	GAATCTACTGTGTTACATCACATGTCTCTGGTCTATTAACA	1233	
QY	1141	TTTGAAACAGACCCCTG	AGAGATCAACACATATGACGGCTGATCTTCAAAAGATTTAAAGCTT	1200	
Db	1234	TTTGAAACAGACCCCTG	AGAGATCAACACATATGACGGCTGATCTTCAAAAGATTTAAAGCTT	1293	
QY	1201	AGTAGAGACTGTGCA	CAATCAATCGTCAAGGTGTGCAAGCTCAATGAATGGGAGATATGTCC	1266	
Db	1294	AGTAGAGACTGTGCA	CAATCAATCGTCAAGGTGTGCAAGCTCAATGAATGGGAGATATGTCC	1353	
QY	1261	CATGGCGCTACACTT	ATAGGAGAAATATACCATGTGATCATGATGATGCTCATGACCCCAAA	1320	
Db	1354	CATGGCGCTACACTT	ATAGGAGAAATATACCATGTGATCATGATGATGCTCATGACCCCAAA	1413	
QY	1321	AGACCTGGAGACTTGG	ATGATGTTATATGAGATATCCCATGGAGATGTTCAATCAACTAGA	1388	
Db	1414	AGACCTGGAGACTTGG	ATGATGTTATATGAGATATCCCATGGAGATGTTCAATCAACTAGA	1477	
QY	1381	ACCTTTGATCATCTTA	TAAATGTAAGATCTTAAAGCAAGCTTCCCTTGATATATGACAGGG	1444	
Db	1474	ACCTTTGATCATCTTA	TAAATGTAAGATCTTAAAGCAAGCTTCCCTTGATATATGACAGGG	1533	
QY	1441	GGACATGAGTGAAGTGA	AGAGGTTGCTTCTGTGTTATAGACAGGGCA	1491	
Db	1534	GGACATGAGTGAAGTGA	AGAGGTTGCTTCTGTGTTATAGACAGGGCA	1584	

```

RESULT 2      ACI04634/c    115758 bp   DNA     linear PRI 10-AUG-2002
LOCUS         ACI04634          115758 bp   DNA     linear PRI 10-AUG-2002
DEFINITION    Homo sapiens BAC clone RP11-793L24 from 2, complete sequence.
ACCESSION     ACI04634
VERSION       ACI04634.5 GI:22002223
KEYWORDS      HTG.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE     Mammalian Genome Project International Consortium, et al., Nature Genet. 1(1), 11-16 (1993).
AUTHORS       Sulston, J.E. and Waterston, R.
TITLE         Toward a complete human genome sequence
JOURNAL       Genome Res. 8 (11), 1057-1108 (1998)
MEDLINE       99063792
REFERENCE     Du, H., Kozlowicz, A. and Spalding, L.
AUTHORS       The Human Genome Mapping International Working Group, et al., Science 261(5128), 116-121 (1993).
TITLE         Unpublished (2001)
JOURNAL       3 (bases 1 to 115758)
REFERENCE     Waterston, R.H.
AUTHORS       Direct Submission
TITLE         Submitted (15-DEC-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
JOURNAL       4 (bases 1 to 115758)
REFERENCE     Waterston, R.H.
AUTHORS       Direct Submission
TITLE         Submitted (21-FEB-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
JOURNAL       5 (bases 1 to 115758)
REFERENCE     Waterston, R.H.
AUTHORS       Direct Submission
TITLE         Submitted (29-JUL-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
JOURNAL       6 (bases 1 to 115758)
REFERENCE     Waterston, R.
AUTHORS       Direct Submission
TITLE         Submitted (10-AUG-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
JOURNAL       On Jul 29, 2002 this sequence version replaced gi:18855170.
COMMENT       --- Genome Center
               Center: Washington University Genome Sequencing Center
               Center code: WUGSC
               Web site: http://genome.wustl.edu/gsc
               Contact: saplens@wustl.wustl.edu
               Summary Statistics
               Center project name: H_NH0793L24
               -----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry or covered by high quality data [i.e., phred quality >-
30]; an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis,
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

```

```

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P. Y., Zhao, B., Frengen, E., Tatemio, M., Caranese, J. J. and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-56215, 2000 bp overlap. The clone sequenced to the right is RP11-690121, 2000 bp overlap. Actual start of this clone is at base position 131499 of RP11-56215; actual end is at base position 65545 of RP11-690121.

Simple sequence repeats exists between 20489 and 21499.
Polymorphisms exist between AC017104, AC073476 and AC104634.

FEATURES
source
1..115758
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-793L24"
/clone_1lb="RPCI-11"
1..280
/rpt_family="Alu"
289..320
/rpt_family="(TAA)n"
335..634
/rpt_family="Alu"
1043..1348
/rpt_family="Alu"
1356..1489
/rpt_family="Alu"
2044..2092
/rpt_family="(TAA)n"
2094..2424
/rpt_family="Alu"
2427..2605
/rpt_family="Alu"
2609..2676
/rpt_family="WAlR"
2759..2790
/rpt_family="AT-rich"
2791..3080
/rpt_family="Alu"
3083..3378
/rpt_family="Alu"
3441..3564
/rpt_family="L1"
3602..3643
/rpt_family="Alu"
3645..3838
/rpt_family="Alu"
3899..4051
/rpt_family="L1"
4300..4589
/rpt_family="Alu"
4686..4846
/rpt_family="L1"
5660..5935
/rpt_family="Alu"
6048..6347
/rpt_family="Alu"
6383..6447
/rpt_family="WAlR"
6448..6737
/rpt_family="Alu"
6743..6778
/rpt_family="(TTA)n"
6779..6807

```

```
repeat_region /rpl_family="MaLR"
6808. 6843 /rpl_family="ERV1"
repeat_region /rpl_family="Alu"
6844. 7149 /rpl_family="Alu"
7150. 7382 /rpl_family="ERV1"
repeat_region /rpl_family="Alu"
7383. 7681 /rpl_family="Alu"
repeat_region /rpl_family="ERV1"
7682. 7913 /rpl_family="ERV1"
7914. 8310 /rpl_family="MaLR"
repeat_region /rpl_family="Alu"
8315. 8625 /rpl_family="Alu"
8620. 9199 /rpl_family="MaLR"
repeat_region /rpl_family="MaLR"
9205. 9366 /rpl_family="MER103"
repeat_region /rpl_family="Alu"
9371. 9686 /rpl_family="Alu"
repeat_region /rpl_family="MaLR"
9703. 9875 /rpl_family="Alu"
repeat_region /rpl_family="Alu"
11178. 11536 /rpl_family="ERV1"
repeat_region /rpl_family="ERV1"
11839. 11880 /rpl_family="ERV1"
repeat_region /rpl_family="Alu"
11881. 12178 /rpl_family="Alu"
repeat_region /rpl_family="Alu"
12179. 12419 /rpl_family="ERV1"
repeat_region /rpl_family="ERV1"
12420. 12722 /rpl_family="Alu"
repeat_region /rpl_family="Alu"
12877. 13173 /rpl_family="Alu"
repeat_region /rpl_family="ERV1"
14200. 14389 /rpl_family="Al"
repeat_region /rpl_family="Al"
14403. 14429 /rpl_family="(T)n"
repeat_region /rpl_family="(T)n"
14466. 14776 /rpl_family="Alu"
repeat_region /rpl_family="Alu"
14825. 14878 /rpl_family="(TC)n"
repeat_region /rpl_family="MIR"
16155. 16213 /rpl_family="Alu"
repeat_region /rpl_family="Alu"
16214. 16523 /rpl_family="Alu"
repeat_region /rpl_family="MaLR"
16625. 16848 /rpl_family="Alu"
repeat_region /rpl_family="Alu"
16854. 17157 /rpl_family="Alu"
repeat_region /rpl_family="MaLR"
17165. 17277 /rpl_family="MaLR"
```

Query Match 5.9%; Score 88.4; DB 9; Length 115758;
Best Local Similarity 43.4%; P-1e-11;
Matches 398; Conservative 1; Mismatches 517; Indels 0; Gaps 0;

```
QY 1 GAVGGGAGCAGTGTACGATGGGCAAGCAAAAGATGACCAACCATGACCAACCAACGAT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 21399 GATGGTGTGGTGAAGTGTACTGTATGTATGTATGTATGTATGTATGTATGTATGTAT 21340
QY 61 GATACACATGACGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 21339 GATGGTGTATGCTGTATGCTGTATGCTGTATGCTGTATGCTGTATGCTGTATGCTGTAT 21280
QY 121 GACCAACCAACCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 21279 AATGATGTGTATGCTGTATGCTGTATGCTGTATGCTGTATGCTGTATGCTGTATGCTGT 21220
```

```
QY 181 TCACAGACGGTCATGAGACCTGTTATCTAGAACCTCATCTGCGATTCAACCAAGT 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 21219 AATATTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 21160
QY 241 GAAGACCATGACGACCAACCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 21159 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 21100
QY 301 GGTGTTGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 21099 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 21040
QY 361 GACCTGGTGACCTGGTTGACATGATAGGGCGCTGTTAATGAATTCATCATATGCT 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 21039 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 20980
QY 421 TGGTTGACATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 20979 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 20920
QY 481 ATTTTCAAGGAGTCACACCGATGCTGATACCCAGCAAGTAATCCCTGTTGCTT 540
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 20919 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 20860
QY 541 ATTGTCATGAGAAAGCTGCGCCAGAACACGACCTGCTACATCAGATGAGAGAA 600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 20859 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 20800
QY 601 GATTAACAGGACATTAATGCCATTTGACCTAGATCTAATACACCAACCAAGGCT 660
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 20799 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 20740
QY 661 CTCATCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 20739 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 20680
QY 721 GTGTCTACCATTTAGAGGATTTAATAGTATGATGATGATGATGATGATGATGATGATGAT 780
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 20679 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 20620
QY 781 GTACAGATCTACGCCACGCTGACCTGACAGCTGATGTATTAACCTCGGTCTAATAT 840
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 20619 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 20560
QY 841 GATCCTCATGATTAACCATTAACCATGATGATGATGATGATGATGATGATGATGATGATGAT 900
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 20559 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 20500
QY 901 CATGGCTTGTCTAATG 916
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 20499 GATGGTATGGGGATG 20484
```

RESULT 3
AC079176 186278 bp DNA linear PRI 29-JUN-2002
LOCUS Homo sapiens X BAC Rpl1-325Ds (Roswell Park Cancer Institute Human
DEFINITION BAC library) complete sequence.
AC079176 AC079176 GI:21629127
VERSION AC079176.15
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 186278)
Muzny,D.M., Adams,C., Adio-Obuola,B., All-ouman,F.R., Allen,C.,
Alstbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbarella,J.,
Beneton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brievara,M., Brown,E., Brown,M., Bryant,N.P., Bulay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Cartron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,

Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Dem, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Drepper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, J., Garza, N., Gill, R., Correll, J.H., Guevara, W., Gunnarsson, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Hayak, P., Hayes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Hult, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, T., Johnson, R., Jollivet, S., Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Koryah, J., Kovar, C., Kravovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichteage, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., Moleod, M.P., Meador, M., Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokewo, S., Ogutu, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Ruit, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shostkari, N., Sisson, I., Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Swalek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vazquez, L., Vera, V., Villalobos, C., Winson, R., Wang, O., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Washington, S., Williams, G., Williamson, A., Wleczek, R., Woodson, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 186278)
Worley, K.C.

Direct Submission
Submitted (23-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One 3 (bases 1 to 186278)
Worley, K.C.

Direct Submission
Submitted (29-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jun 29, 2002 this sequence version replaced gi:1326484.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2

reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: http://www.hgsc.bcm.tmc.edu:8088/quality_info/genbank.annot

FEATURES

source	Location/Qualifiers
repeat_region	1..186278 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="X" /clone="RP11-325D5"
repeat_region	195..244 /rpt_family="(CA)n"
repeat_region	1562..1592 /rpt_family="(GA)n"
repeat_region	1862..1939 /rpt_family="(GC-rich)"
repeat_region	2060..2135 /rpt_family="(GC-rich)"
repeat_region	2293..2389 /rpt_family="(GC-rich)"
repeat_region	25328..5638 /rpt_family="(A)usg"
repeat_region	53283..60555 /rpt_family="(A)usg"
repeat_region	6132..6163 /rpt_family="(TG)n"
repeat_region	6344..6647 /rpt_family="(A)usx"
repeat_region	7051..7173 /rpt_family="(FLAM-C)"
repeat_region	7188..7482 /rpt_family="(A)usx"
repeat_region	9120..9424 /rpt_family="(A)usx"
repeat_region	9835..10137 /rpt_family="(A)usx"
repeat_region	10355..10578 /rpt_family="(A)usg"
repeat_region	10583..10628 /rpt_family="(A)usg"
repeat_region	10629..10822 /rpt_family="(A)usg"
repeat_region	10852..11168 /rpt_family="(A)usx"
repeat_region	11349..11473 /rpt_family="(MER3)"
repeat_region	11512..11741 /rpt_family="(MT1B)"
repeat_region	11742..12124 /rpt_family="(MT1B)"
repeat_region	12505..12536 /rpt_family="(CG)n"
repeat_region	12536..12568 /rpt_family="(CA)n"
repeat_region	12672..12961 /rpt_family="(A)usg"
repeat_region	13662..13970 /rpt_family="(A)usg"
repeat_region	14644..14944 /rpt_family="(A)usx"
repeat_region	15076..15137 /rpt_family="(TA)n"
repeat_region	15232..15407 /rpt_family="(TA)n"

```

repeat_region 15427. .15485 /rpt_family="AT-rich"
repeat_region 15513. .15598 /rpt_family="(TA)n"
repeat_region 15599. .15892 /rpt_family="Alusp"
repeat_region 15688. .15803 /standard_name="D11S2560"
STS 15893. .15979 /rpt_family="(TA)n"
repeat_region 16010. .16170 /rpt_family="(TA)n"
repeat_region 16224. .16405 /rpt_family="MER33"
repeat_region 16406. .16832 /rpt_family="MSFB"
repeat_region 16833. .16881 /rpt_family="MER33"
repeat_region 17271. .17567 /rpt_family="Aluy"
repeat_region 17528. .17781 /rpt_family="L1MC3"
repeat_region 17816. .17866 /rpt_family="MSFB"
repeat_region 17951. .18065 /rpt_family="L1MC3"
repeat_region 18066. .18752 /rpt_family="L1MC3"
repeat_region 18787. .19078 /rpt_family="Aluuo"
repeat_region 19082. .19165 /standard_name="STS3076"
STS 19082. .19165 /standard_name="STS3076"
repeat_region 20590. .20723 /rpt_family="(TTA)n"
repeat_region 20724. .21004 /rpt_family="(TTA)n"

```

[illegible][illegible]

LOCUS	AC009053.3	RESULT 4
DEFINITION	Mus musculus clone RP23-33H13 strain C57BL6/J, WORKING DRAFT	
ACCESSION	AC009053.3	
VERSION	AC009053.3.4	
KEYWORDS	HTG: HTGS_PHASE1, HTGS_DRAFT.	
SOURCE	Mus musculus.	
ORGANISM	Mus musculus.	
REFERENCE	Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS	Li, L., Montgomery, K.T., Gillis, G., Chiu, D., Decker, J., Fusina, M., 1 (bases 1 to 272545) Goltz, J., Halder, A., Hall, L., Han, J., Ioshikhes, I. P., Lee, E., Perera, A., Shum, C., Thomas, E. and Kucherlapati, R. High Throughput Mouse Sequencing Unpublished 2 (bases 1 to 272545) Li, L., Montgomery, K.T., Gillis, G., Chiu, D., Decker, J., Fusina, M., Goltz, J., Halder, A., Hall, L., Han, J., Ioshikhes, I. P., Lee, E., Perera, A., Shum, C., Thomas, E. and Kucherlapati, R.	
TITLE		
JOURNAL		
REFERENCE		
AUTHORS		
COMMENT	Center: Harvard Partners Genome Center Center: HPGC Web site: http://www.hpcg.org/Sequence/mouse.html Contact: hpgc@emdel.mgh.harvard.edu -----Summary Statistics Center project name: ADU Sequencing vector: pUC18, 108752 Chemistry: Dye-terminator Big Dye, 100% *Consensus quality: 257441 at least Q20 *Consensus quality: 253718 at least Q30 *Consensus quality: 247095 at least Q40	

misc_feature /note="assembly_name:Contig92"
230383..236251
/note="assembly_name:Contig91"
236272..240572
misc_feature /note="assembly_name:Contig90"
240593..242789
misc_feature /note="assembly_name:Contig89"
242810..244695
misc_feature /note="assembly_name:Contig88"
244716..246969
misc_feature /note="assembly_name:Contig87"
246990..247870
misc_feature /note="assembly_name:Contig86"
247891..248193
misc_feature /note="assembly_name:Contig85"
248214..250122
/note="assembly_name:Contig84"

Query Match 4.7% Score 69.6; DB 2; Length 272545;
Best Local Similarity 49.3%; Pred. No. 9.6e-07;
Matches 206; Conservative 1; Mismatches 210; Indels 1; Gaps 1;

Qy 1 GAYGGGAGAGCTGTAACGATGCGGACAGACAGATGACACACAGACAGACAGATG 60
Db 251726 GATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 251785
Qy 61 GATCAGCATGACGACATGATGATGATGATGATGATGATGATGATGATGATGATG 120
Db 251786 CATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 251845
Qy 121 GAACCAACCCCTAATGAGGCTGAGAGCCCTTACACACATGCTAGGAGCATAGCTG 180
Db 251846 CATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 251905
Qy 181 TCACGAAGGCTGATGAGGCTTATCTAGAACTTCTGTGGATCAACCAAGT 240
Db 251906 GATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 251965
Qy 241 GAAGACATGACGACACCATCATCATCATCATCATCATCATCATCATCATCATCAT 229
Db 251966 CATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 252025
Qy 300 AGCTTGATCTTATGCGCAACTGTACAACTGCTACCCAGAAAAACATGCTGACCTGG 359
Db 252026 TCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 252085
Qy 360 TCACCTCGTACCTGCTTGCATGATGATGATGATGATGATGATGATGATGATGAT 417
Db 252086 TCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 252143

RESULT 5
AL627248/c 143716 bp DNA linear HTG 12-AUG-2002
LOCUS Dario rerio clone RP71-1618, *** SEQUENCING IN PROGRESS ***, 2
DEFINITION unordered pieces.
AL627248
VERSION AL627248.5 GI:22204283
KEYWORDS HTG; HTGS; PHASE1; HTGS_ACTIVEFTN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE zebrafish.
ORGANISM Dario rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 143716)
Babbage/A.
Direct Submission
Submitted (08-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zface@sanger.ac.uk
On Aug 11, 2002 this sequence version replaced gi:22003110.
----- Genome Center
Center: Wellcome Trust Sanger Institute

Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zface@sanger.ac.uk
----- Project Information
Center project name: bz1818
----- Summary Statistics
Sequencing vector: XGAP4; version 4.5
Chemistry: Dye-terminator, 1% of reads
Chemistry: Dye-terminator, Big Dye, 98% of reads
Consensus quality: 143560 bases at least Q40
Consensus quality: 143567 bases at least Q30
Consensus quality: 143614 bases at least Q20
Insert size: 152033; 4.8% error; agarose-fp
Quality coverage: 9.51x in Q20 bases; sum-of-contigs Quality
coverage: 8.98x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 86951: contig of 86951 bp in length
* 86952 87051: gap of 100 bp
* 87052 143716: contig of 56665 bp in length.
FEATURES
location/Qualifiers
source 1
1..143716
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="RP71-1618"
/clone_1db="RPCT-71"
1..86951
/note="assembly_fragment:00443
fragment_chain:1"
87052..143716
/note="assembly_fragment:00635
fragment_chain:1"

BASE COUNT 47522 a 25963 c 25217 g 44914 t 100 others
ORIGIN
Query Match 4.5% Score 66.4; DB 2; Length 143716;
Best Local Similarity 46.8%; Pred. No. 6.1e-06;
Matches 205; Conservative 1; Mismatches 232; Indels 0; Gaps 0;

Qy 1 GAYGGGAGAGCTGTAACGATGCGGACAGACAGATGACACACAGACAGATG 60
Db 8441 GATGACGACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8382
Qy 61 GATACCATGACGACGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
Db 8381 GATGACGATGACGACGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 8322
Qy 121 GAACCAACCCCTAATGAGGCTTACACACATGCTGACATGACGACAGATGATG 180
Db 8321 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8262
Qy 181 TCACGAAGGCTGATGAGCTGTTATCTAGAACTTCTGTGGATCAACCAAGT 240
Db 8261 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8202
Qy 241 GAAGACATGACGACGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
Db 8201 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8142
Qy 301 GCTTGATCTTATGCGCAACTGCTCAATGCTCAACCAACCAATGCTGACCTGGT 360
Db 8141 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8082
Qy 361 GACCTCGTACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420

|||||
Db 8081 GAGCAGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8022
Oy 421 TGCTTGACATGATGATGAT 438
Db 8021 GATGACGATGATGATGATGAT 8004

RESULT 6
AC096302
LOCUS
DEFINITION
AC096302 222546 bp DNA linear HMG 11-JUL-2002
Rattus norvegicus clone CH230-99P5, *** SEQUENCING IN PROGRESS ***,
67 unordered pieces.
AC096302
VERSION
AC096302.5 GI:21723460
HTG: HTGS_PHASE1.
KEYWORDS
Norway rat.
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS
1 (bases 1 to 222546)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alshrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbata,J., Benton,J., Bimarge,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrnes,N.C.,
Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.T.,
Earmhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frintz,P.,
Gabrelli,J.H., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hale,S., Hamilton,K.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homs,J.F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolyvet,S., Joudah,S.,
Karlovic,J., Kureshi,A., Landry,N., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,R., Louisedge,H.,
Lozdo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,E., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogun,M., Okwunonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Saverly,G.,
Sodergren,E., Soneike,T., Sparks,A., Stanley,H., Stone,K.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vilson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wlcek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
TITLE
Submitted (17-Sep-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 222546)
Worley,K.C.
Direct Submission
Direct Submission

JOURNAL
COMMENT

Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 10, 2002 this sequence version replaced gi:20976083.

Center: Baylor College of Medicine
Genome Center
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project information
Center project name: GEWV
Center clone name: CH230-99P5

----- Summary Statistics -----
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 159115 bases at least Q40
Consensus quality: 169276 bases at least Q30
Consensus quality: 176785 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 67 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1110: contig of 1110 bp in length
* 1111 1210: gap of unknown length
* 1211 2450: contig of 1240 bp in length
* 2451 2550: gap of unknown length
* 2551 3788: contig of 1238 bp in length
* 3789 3889: gap of unknown length
* 3889 5437: contig of 1549 bp in length
* 5438 5537: gap of unknown length
* 5537 6906: contig of 1369 bp in length
* 6907 7006: gap of unknown length
* 7007 8334: contig of 1328 bp in length
* 8335 8435: gap of unknown length
* 8435 9601: contig of 1167 bp in length
* 9602 9701: gap of unknown length
* 9701 11677: contig of 1976 bp in length
* 11678 11777: gap of unknown length
* 11778 13136: contig of 1359 bp in length
* 13137 13236: gap of unknown length
* 13237 14323: contig of 1087 bp in length
* 14324 14423: gap of unknown length
* 14423 15448: contig of 1325 bp in length
* 15449 15848: gap of unknown length
* 15849 17323: contig of 1475 bp in length
* 17324 17423: gap of unknown length
* 17424 18648: contig of 1225 bp in length
* 18649 18748: gap of unknown length
* 18749 20689: contig of 1941 bp in length
* 20690 20789: gap of unknown length
* 20790 24016: contig of 3227 bp in length
* 24017 24116: gap of unknown length
* 24117 25552: contig of 1436 bp in length
* 25553 25652: gap of unknown length
* 25653 27378: contig of 1726 bp in length
* 27379 27478: gap of unknown length
* 27479 30784: contig of 3306 bp in length
* 30785 30884: gap of unknown length
* 30885 32845: contig of 2061 bp in length
* 32846 33045: gap of unknown length
* 33046 33143: contig of 2098 bp in length
* 33144 35243: gap of unknown length
* 35244 36966: contig of 1723 bp in length
* 36967 37067: gap of unknown length
* 37068 38526: contig of 1460 bp in length

```

38527 38626: gap of unknown length
38627 40316: contig of 1690 bp in length
40317 40316: gap of unknown length
40417 42559: contig of 2143 bp in length
42560 42559: gap of unknown length
42660 44169: contig of 1510 bp in length
44170 44270: gap of unknown length
44270 46645: contig of 2376 bp in length
46646 47790: gap of unknown length
46746 47790: contig of 1045 bp in length
47791 47891: gap of unknown length
47891 49740: gap of unknown length
49740 51171: gap of unknown length
51171 53978: contig of 2708 bp in length
53978 54078: gap of unknown length
54078 55262: contig of 1184 bp in length
55263 55362: gap of unknown length
55363 57537: contig of 2075 bp in length
57538 59168: gap of unknown length
59169 59268: contig of 1631 bp in length
59269 61614: contig of 2346 bp in length
61615 61715: gap of unknown length
61715 64695: contig of 2981 bp in length
64696 64795: gap of unknown length
64796 67778: contig of 2883 bp in length
67779 70112: gap of unknown length
70113 70212: contig of 2334 bp in length
70213 73804: gap of unknown length
73805 73904: contig of 3592 bp in length
73905 76316: gap of unknown length
76317 76416: contig of 2412 bp in length
76417 79376: gap of unknown length
79377 79476: contig of 2960 bp in length
79477 82449: gap of unknown length
82450 82349: contig of 2773 bp in length
82350 84713: gap of unknown length
84713 84812: contig of 2363 bp in length
84812 87975: gap of unknown length
87976 88075: contig of 3163 bp in length
88076 91512: gap of unknown length
91513 91612: contig of 3437 bp in length
91613 95939: gap of unknown length
95940 96039: contig of 4327 bp in length
96040 98464: gap of unknown length
98465 98564: gap of unknown length
98565 103056: contig of 2425 bp in length
103057 103156: gap of unknown length
103157 103256: contig of 4492 bp in length
103257 107561: gap of unknown length
107562 107661: contig of 4405 bp in length
107662 111867: gap of unknown length
111868 111967: contig of 4206 bp in length
111968 116755: gap of unknown length
116756 116855: contig of 4788 bp in length
116856 122068: gap of unknown length
122069 122068: contig of 5113 bp in length
122069 127578: gap of unknown length
127579 127578: contig of 5510 bp in length
127679 131347: gap of unknown length
131348 131447: contig of 3669 bp in length
131448 140101: gap of unknown length
140102: contig of 8654 bp in length

```

Query Match 4.4% Score 65.2; DB 2: Length 222546;
 Best Local Similarity 47.0%; Pred. No. 1.3e-05;
 Matches 196; Conservative 1; Mismatches 220; Indels 0; Gaps 0;

OY 1 GAYGGGACGCTGTACGATGGCGACAAAGATGACACCGACGACCGACCGAT 60
 11:1 11:1 11:1 11:1 11:1 11:1 11:1 11:1 11:1 11:1
 Db 33 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 92

```

OY 61 GATCACCATGACGACCATGATGATGATGAACAATGACCTATGCCAGTGGAATG 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 93 CATGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 152
OY 121 GAACCAAAACCTCAATGAGCTAGACGCTTACACCATGCTGCAAGCATGAGT 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 153 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 212
OY 181 TCACGAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 213 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 272
OY 241 GAAGACATGACGACCATGATGATGATGATGATGATGATGATGATGATGAT 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 273 CACCATCATGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 332
OY 301 GGTGTGATCTTATGCGAAGCTGACATGCTACCAAGAAACATGCTGACCT 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 333 GATGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 392
OY 361 GACCTGCTGACCTGCTGATGATGATGATGATGATGATGATGATGATGAT 417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 393 GACGACGATCATGCGCATGACACCGGACGATGATGATGATGATGATGAT 449

```

RESULT 7
 AC090533/c 272545 bp DNA linear HTG 15-MAY-2002
 LOCUS Mus musculus clone RP23-33H13 strain C57BL6/J, WORKING DRAFT
 DEFINITION
 AC090533
 AC090533.4 GI:17976439
 VERSION
 AC090533
 KEYWORDS
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 SOURCE
 Mus musculus.
 ORGANISM
 Mus musculus

REFERENCE
 1 (bases 1 to 272545)
 I.L., Montgomery,K.T., Grills,G., Chiu,D., Decker,J., Fusina,M.,
 Goltz,J., Halder,A., Hall,L., Han,J., Joshikhes,I.P., Lee,E.,
 Pereira,A., Shim,C., Thomas,E. and Kucherlapati,R.
 High Throughput Mouse Sequencing

TITLE
 JOURNAL
 REFERENCE
 AUTHORS

TITLE
 JOURNAL
 REFERENCE
 AUTHORS

COMMENT
 On Dec 21, 2001 this sequence version replaced g1:13310871.

Center: Harvard Partners Genome Center
 Center Code: HPGC
 Web site: <http://www.hpcg.org/Sequence/mouse.html>
 Contact: hpcg@emdel.mgh.harvard.edu

Center project name: ADU
 Sequencing vector: pUC18; L08752
 Chemistry: Dye-terminator Big Dye; 100#
 *Consensus quality: 257441 at least Q20
 *Consensus quality: 253718 at least Q30
 *Consensus quality: 247095 at least Q40
 *Estimated insert size: agarose-pp - N/A
 *Estimated insert size: 271665 - sum-of-contigs
 Quality coverage: 6.8 x in Q20 bases: sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 45 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 30267: contig of 30267 bp in length
* 30268 30287: gap of unknown length
* 30288 58400: contig of 2813 bp in length
* 58401 58420: gap of unknown length
* 58421 88864: contig of 3044 bp in length
* 88865 88884: gap of unknown length
* 88885 109890: contig of 21005 bp in length
* 109891 109909: gap of unknown length
* 109910 125133: contig of 15224 bp in length
* 125134 125153: gap of unknown length
* 125154 139549: contig of 14396 bp in length
* 139550 139569: gap of unknown length
* 139570 153410: contig of 13841 bp in length
* 153411 153431: gap of unknown length
* 153432 162567: contig of 9137 bp in length
* 162568 162587: gap of unknown length
* 162588 171044: contig of 8457 bp in length
* 171045 171065 179304: contig of 8240 bp in length
* 179305 179324: gap of unknown length
* 179325 187059: contig of 7735 bp in length
* 187060 187079: gap of unknown length
* 187080 197598: contig of 10519 bp in length
* 197599 197618: gap of unknown length
* 197619 205986: contig of 8348 bp in length
* 205987 205986: gap of unknown length
* 205987 212449: contig of 6463 bp in length
* 212450 212469: gap of unknown length
* 212470 217752: contig of 5283 bp in length
* 217753 217772: gap of unknown length
* 217773 223943: contig of 6171 bp in length
* 223944 223963: gap of unknown length
* 223964 230362: contig of 6399 bp in length
* 230363 230382: gap of unknown length
* 230383 236251: contig of 5869 bp in length
* 236252 236271: gap of unknown length
* 236272 240572: contig of 4301 bp in length
* 240573 240592: gap of unknown length
* 240593 242789: contig of 2197 bp in length
* 242790 242809: gap of unknown length
* 244695 244715: gap of unknown length
* 244716 246969: contig of 1886 bp in length
* 246970 246989: gap of unknown length
* 246990 247870: contig of 2254 bp in length
* 247871 247890: gap of unknown length
* 247891 248193: contig of 881 bp in length
* 248194 248193: gap of 303 bp in length
* 248194 248213: gap of unknown length
* 248214 250122: contig of 1909 bp in length
* 250123 250142: gap of unknown length
* 250143 251367: contig of 1225 bp in length
* 251368 251387: gap of unknown length
* 251388 252749: contig of 1362 bp in length
* 252750 252769: gap of unknown length
* 252770 253640: contig of 871 bp in length
* 253641 253660: gap of unknown length
* 253661 255032: contig of 1372 bp in length
* 255033 255052: gap of unknown length
* 255053 257333: contig of 2281 bp in length
* 257334 257353: gap of unknown length
* 257354 259409: contig of 2056 bp in length
* 259410 259429: gap of unknown length
* 259430 260459: contig of 1030 bp in length
* 260460 260479: gap of unknown length
* 260480 260829: contig of 350 bp in length
* 260830 260849: gap of unknown length
* 260850 261651: contig of 812 bp in length
* 261652 261669: gap of unknown length
* 261682 262532: contig of 851 bp in length

262533 262552: gap of unknown length
* 262553 263351: contig of 799 bp in length
* 263352 263371: gap of unknown length
* 263372 264217: contig of 846 bp in length
* 264218 264237: gap of unknown length
* 264238 265205: contig of 968 bp in length
* 265206 265225: gap of unknown length
* 265226 266270: contig of 1045 bp in length
* 266271 266290: gap of unknown length
* 266291 266749: contig of 439 bp in length
* 266750 266769: gap of unknown length
* 266770 267718: contig of 949 bp in length
* 267719 267738: gap of unknown length
* 267739 268576: contig of 838 bp in length
* 268577 268596: gap of unknown length
* 268597 270010: contig of 1414 bp in length
* 270011 270030: gap of unknown length
* 270031 271654: contig of 1624 bp in length
* 271655 271675: gap of unknown length
* 271675 272545: contig of 871 bp in length.

Location/Qualifiers
1. .272545
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RP23-33H13"
/sex="male"
1. .30267
/note="assembly_name:Contig108"
30288. .58400
/note="assembly_name:Contig107
clone_end:SP6
vector_side:left"
58421. .88864
/note="assembly_name:Contig106"
88885. .109889
/note="assembly_name:Contig105"
109910. .125133
/note="assembly_name:Contig104
clone_end:T7
vector_side:right"
125134. .139549
/note="assembly_name:Contig103"
139570. .153410
/note="assembly_name:Contig102"
153431. .162567
/note="assembly_name:Contig101"
162588. .171044
/note="assembly_name:Contig100"
171065. .179304
/note="assembly_name:Contig99"
179325. .187059
/note="assembly_name:Contig98"
187080. .197598
/note="assembly_name:Contig97"
197619. .205966
/note="assembly_name:Contig96"
205987. .212449
/note="assembly_name:Contig95"
212470. .217752
/note="assembly_name:Contig94"
217773. .223943
/note="assembly_name:Contig93"
223964. .230362
/note="assembly_name:Contig92"
230383. .236251
/note="assembly_name:Contig91"
236272. .240572
/note="assembly_name:Contig90"
240593. .242789
/note="assembly_name:Contig89"
242810. .244695
/note="assembly_name:Contig88"

	D	b	157652	ATGATGTGGTACTCTTTATTTGGTGATGCATCATATGATGATGAGTCGATGAATTATGTTA	157593
Oy	479	CTAT-----TTACAAAGGAGTCACACCAGATGCGATACCCAGCAGTAGAATCG	529		
D	157592	ATAATGATGGTGATTTATGTATGACTAAGATTAACGATTAATGATGATGGTATGACGATG	157593		
Oy	530	CCGTGTTGTTATTTGGTCATGGAAGAAGCGGCCCAAGAAACGACGCGCTCTACATCACG	589		
D	157532	ATGATGGTATGACGATGATTAATAGTATGGTATGATGATGAAGAATATGATGTCAGTATG	157473		
Oy	590	ACGTAAGGGAAGCTAAACCTGAGCATTTATGCCATTGTGACGTAAATCTAATACACACC	649		
D	157472	AGGTTGATCTAATGATGGTATGATGACGATATGATGATGATGACAAATGAAATGATTAATC	157413		
Oy	650	AACCAAAGCCTCTTCATCATCTGTCACGGAACCCATGATTTCAAAACAAGTTGTTATG	709		
D	157412	ATGGCAATGATGGTGATGATGACAAATATGATNGTGATGATGATGACAAAGATGATGATG	157353		
Oy	710	GTCACCTGTAAGTGCTCTACCACTTTAGAGGGATTTTAATGTAACTGATGACACACAAATGC	769		
D	157352	GTCGTGATTAATGTTGATTAATGATGAAGGTGATTAATGATTAATGAAGATGATGATGAAGATG	157293		
Oy	770	ATCCATCGACGACGATCTACGCCAACGGGACCTCAGTCAGTCAGTATGATTAACCTCG	829		
D	157292	ATGCTATGCTGTTGATGACATATGATTAATGGGAGTATGATGACAGTATGATGATGATGATG	157233		
Oy	830	GTCGTAATATGATTCCTCATGAAGATTTACCAAGTATGGTGGGATCTAGAGATATTC	889		
D	157232	ATGTTTATACCATATGATGATGATGATTAACAAATGATTAATGATGATGATGATGATGATG	157173		
Oy	890	ACGATGATGACCATGGCGCTTGCATCAATGA	919		
D	157172	ACTATGATGATGGTATGATTAATTAATA	157143		
RESULT_9	ACOL18695/c	211001 bp	DNA linear	PRI 09-JAN-2002	
LOCUS	Homo sapiens BAC clone Rpl1-368J23 from 16, complete sequence.				
DEFINITION	ACOL18695				
ACCESSION	ACOL18695.6	GI:13443271			
VERSION					
KEYWORDS	HTG.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 211001)				
AUTHORS	Sulston,J.E. and Waterston,R.				
TITLE	Toward a complete human genome sequence				
JOURNAL	Genome Res. 8 (11), 1097-1108 (1998)				
MEDLINE	99063792				
PUBMED	9847074				
REFERENCE	2 (bases 1 to 211001)				
AUTHORS	Haakenson,W., Scott,K., Courtney,L., Drone,K., Gregory,S. and				
	Lesley,K.				
TITLE	The sequence of Homo sapiens BAC clone Rpl1-368J23				
JOURNAL	Unpublished (2001)				
AUTHORS	3 (bases 1 to 211001)				
TITLE	Waterston,R.H.				
JOURNAL	Direct Submission				
	Submitted (16-Dec-1999) Genome Sequencing Center, Washington				
	University School of Medicine, 4444 Forest Park Parkway, St. Louis				
	MO 63108, USA				
REFERENCE	4 (bases 1 to 211001)				
AUTHORS	Waterston,R.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (24-MAR-2001) Genome Sequencing Center, Washington				
	University School of Medicine, 4444 Forest Park Parkway, St. Louis				
	MO 63108, USA				
REFERENCE	5 (bases 1 to 211001)				
AUTHORS	Waterston,R.				
TITLE	Direct Submission				

```

JOURNAL      Submitted (07-NOV-2001) Department of Genetics, Washington
REFERENCE    University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
AUTHORS      6 (bases 1 to 211001)
TITLE        Waterston, R.
JOURNAL      Direct Submission
COMMENT       Submitted (09-JAN-2002) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
              On Mar 24, 2001 this sequence version replaced gi:9838280.
              ----- Genome Center
              Center: Washington University Genome Sequencing Center
              Center code: WUGSC
              Web site: http://genome.wustl.edu/gsc
              Contact: saplens@wustl.edu
              ----- Summary Statistics
              Center project name: H_NH0568J23
              -----
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCT-11 human BAC library was made from the blood of one male
donor, as described by Osogawa, K., Moon, P.Y., Zhao, B., Frengen, E.,
Tateno, M., Catalanese, J.J. and de Jong, P.J. (1998) An Improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pletier de Jong
and coworkers at the Roswell Park Cancer Institute
(http://pacpac.med.buffalo.edu)
VECTOR: pBAC3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RP11-542M13. Actual start of
this clone is at base position 1 of RP11-568J23; actual end is at
base position 211001 of RP11-568J23.

H_NH0568J23 contains an imperfect GA run from 143896 to 144293, in
which the exact length is unknown. There are approximately 800
bases missing according to the restriction digests and per data.

H_NH0568J23 contains a single stranded region from 144079 to 144172
which contains low quality data.
      location/Qualifiers
      1..211001
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="16"
      /map="16"
      /clone="RP11-568J23"
      /clone_1lb="RPCT-11"
      234..361
      /rpt_family="Alu"
      407..467
      /rpt_family="MIR"
      760..835
      /rpt_family="MIR"
      891..929
      /rpt_family="MER1_type"

```

```
repeat_region 939..1114 /rpl_family="L1"
repeat_region 1115..1254 /rpl_family="Alu"
repeat_region 1255..1555 /rpl_family="Alu"
repeat_region 1530..1555 /rpl_family="Alu"
repeat_region 1556..1723 /rpl_family="(A)n"
repeat_region 1724..2020 /rpl_family="Alu"
repeat_region 2023..2332 /rpl_family="L1"
repeat_region 2334..2455 /rpl_family="Alu"
repeat_region 2672..2776 /rpl_family="MERL-type"
repeat_region 2793..3021 /rpl_family="L1"
repeat_region 3022..3340 /rpl_family="L1"
repeat_region 3024..3051 /rpl_family="Alu"
repeat_region 3207..3218 /rpl_family="(T)n"
misc_feature 3341..3612 /note="match to EST AA760709 (NID:g2809639) n213c11.s1"
repeat_region 3613..3913 /rpl_family="L1"
repeat_region 3914..4009 /rpl_family="Alu"
repeat_region 4010..4303 /rpl_family="L1"
repeat_region 4291..4349 /rpl_family="Alu"
repeat_region 4367..4677 /rpl_family="AT-rich"
repeat_region 5175..5260 /rpl_family="Alu"
repeat_region 5722..5813 /rpl_family="MERL-type"
repeat_region 5916..6022 /rpl_family="MIR"
repeat_region 6092..6370 /rpl_family="L2"
repeat_region 6374..6523 /rpl_family="Alu"
repeat_region 7188..7617 /rpl_family="Alu"
misc_feature 7234..7630 /note="match to EST AA66253 (NID:g2620866) ac42c06.s1"
misc_feature 7464..7775 /note="match to EST R27299 (NID:g783434) yh40d12.s1"
repeat_region 7609..7690 /note="match to EST T57377 (NID:g659238) yb51a03.r1"
repeat_region 7716..7870 /rpl_family="MIR"
repeat_region 7893..7975 /rpl_family="ERV.L"
repeat_region 8057..8158 /rpl_family="MERL-type"
repeat_region 8334..8424 /rpl_family="ERV.L"
repeat_region 8499..8616 /rpl_family="MIR"
misc_feature 9224..9229 /note="match to EST BE883809 (NID:g10332585)"
repeat_region 9373..9517 /rpl_family="MIR"
misc_feature 9455..10166 /note="match to EST BG385747 (NID:g13278901)"
repeat_region 9577..9698
```

```
repeat_region 9817..9909 /rpl_family="MALR"
repeat_region 9939..10037 /rpl_family="MALR"
repeat_region 10047..10172 /rpl_family="MIR"
repeat_region 10173..10308 /rpl_family="Alu"
repeat_region 10309..10611 /rpl_family="Alu"
repeat_region 10590..10611 /rpl_family="Alu"
repeat_region 10612..10797 /rpl_family="(A)n"
repeat_region 11186..11243 /rpl_family="Alu"
repeat_region 11236..11344 /rpl_family="MALR"
repeat_region 11236..11344 /rpl_family="CT-rich"
```

```
Query Match 4.0%; Score 59.6; DB 9; Length 211001;
Best Local Similarity 43.9%; Pred.No.0.00039;
Matches 257; Conservative 0; Mismatches 329; Indels 0; Gaps 0;

Qy 376 GTTGACGATGATAGGCGCTGTTAATGAAGTTCAATATGCTTGTTGACATTTGAT 435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133803 GATGACAAATGCTGATGCTGCATACATGAGATGAGATGATGATGATGATGAT 133744

Qy 436 GGTACAGACCCAAACACGAGCTCTATTGGACACATCAATGACTATTTTACAAGGAGT 495
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133743 GGTGATGCTGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 133684

Qy 496 CACACCATGCTGATACCCACGACATGAGATGATGATGATGATGATGATGATGATGAT 133624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133683 GATGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 133624

Qy 556 GGTGCCCCAAACAGAGCTGCTTACATCAGAGCTAGAGAGATGATGATGATGATGATGAT 615
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133622 GGTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 133564

Qy 616 TATGCCATTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 675
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133563 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 133504

Qy 676 CACGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 735
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133503 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 133444

Qy 736 GAGGATTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 795
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133443 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 133384

Qy 796 AAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 855
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133383 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 133324

Qy 856 TCCACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 915
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133322 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 133264

Qy 916 GAAGCCACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 961
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133263 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 133218
```

```
RESULT 10
AC117074 61052 bp DNA linear HTG 06-APR-2002
LOCUS Dictyostelium discoideum chromosome 2 map 2779865-2840915 strain
DEFINITION AX4, *** SEQUENCING IN PROGRESS ***; in ordered pieces.
ACCESSION AC117074
VERSION AC117074.1 GI:20066241
KEYWORDS HTG; HTGS-PHASE2.
```

```
SOURCE Dictyostelium discoideum.
ORGANISM Eukaryota; Mycetozoa; Dictyostelida; Dictyostellum.
REFERENCE 1 (bases 1 to 61052)
AUTHORS Gloeckner, G., Eichinger, L., Szafranski, K., Pachbat, J., Dear, P.,
Lehmann, R., Baumgart, C., Parra, G., April, J. F., Guigo, R., Kumpf, K.,
Tunggal, B., Cox, E., Quail, M. A., Platzer, M., Rosenthal, A. and
Neigel, A. A.
TITLE Sequence and Analysis of Chromosome 2 of Dictyostellum
JOURNAL Unpublished
REMARK The Dictyostellum Genome Sequencing Consortium
REFERENCE 2 (bases 1 to 61052)
AUTHORS Baumgart, C.
TITLE Direct Submission
REMARK Submitted (06-APR-2002) Genome Analysis, Institute of Molecular
JOURNAL Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
REFERENCE CDS predictions from Geneid may contain errors. Further information
AUTHORS is available from IMB Jena, Department of Genome Analysis
COMMENT (http://genome.imb-jena.de/dictyostellum/)
and the University Cologne, Institute for Biochemistry I
(http://www.uni-koeln.de/dictyostellum/project.shtml)
Funding
Agency : Deutsche Forschungsgemeinschaft (DFG).
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1. 61052
/organism="Dictyostellum discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/chromosome="2"
/map="2779865-2840915"
BASE COUNT 25445 a 6805 c 6328 g 22474 t
ORIGIN
Query Match 3.9%; Score 58.4; DB 2; Length 61052;
Best Local Similarity 45.7%; Pred. No. 0.00067;
Matches 200; Conservative 1; Mismatches 237; Indels 0; Gaps 0;
FEATURES
source
1 GATGGGAGACAGCTGTAACATGATGAGCAACAAAGATGACACCATGACACCAACAT 60
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7703 GATGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7762
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 61 GATCACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7763 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7822
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 121 GAACCAACCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7823 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7882
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 181 TCACAGAGAGGTCATGAGCTGTTTATCTAGAACTTCATCTGCGATTCAACACAAGT 240
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7883 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7942
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 241 GAAGACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATGAT 300
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7943 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8002
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 301 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8003 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8062
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 361 GACCTCGGAGACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8063 GATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8122
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 421 TCGTTGACATTTGATGAT 438
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8123 TATTAGATTATGAGGAT 8140
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


QY 181 TCACAGAGGTCATGAGACCTGTTATCTAGACTCTGCGATTCAACACAGT 240
 DB 7883 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7942
 QY 241 GAAGACCATGAGACCATGATGATGATGATGATGATGATGATGATGATGAT 300
 DB 7943 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8002
 QY 301 GGTGTCATTCATTCGCGAAGTCTGATGATGATGATGATGATGATGATGAT 360
 DB 8003 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8062
 QY 361 GACCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 420
 DB 8063 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8122
 QY 421 TGGTGGACATTCATGATGATGATGATGATGATGATGATGATGATGATGAT 438
 DB 8123 TATTAGCATTCATGATGATGATGATGATGATGATGATGATGATGATGAT 8140

RESULT 12
 AC131509 163690 bp DNA linear HTG 23-AUG-2002
 LOCUS Strongylocentrotus purpuratus clone sp58L24, *** SEQUENCING IN
 DEFINITION PROGRESS ***, 7 ordered pieces.
 AC131509
 AC131509.1 GI:22450561
 HTG: HTGS_PHASE2, HTGS_FULLTOP.
 STRONGYLOCENTROTUS PURPURATUS.
 STRONGYLOCENTROTUS PURPURATUS.
 Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 Echinoidea; Echinoidea; Echinoidea; Echinoidea;
 Strongylocentrotidae; Strongylocentrotus.
 1 (bases 1 to 163690)
 Davidson, E.H., Rast, J.P., Oliveri, P., Ramsick, A., Castellan, C.,
 Yuh, C.H., Minokawa, T., Amore, G., Himmann, V., Arenas-Mena, C.,
 Schlitz, M.J., Clarke, P.J.C., Rust, A.G., Pan, Z., Arnone, M.I.,
 Rowen, L., Cameron, R., Andrew, M., McClay, D.R., Hood, L. and Bolouri, H.
 A provisional regulatory gene network for specification of
 endomesoderm in the sea urchin embryo
 Dev. Biol. 246 (1), 162-190 (2002)
 22024154
 2 (bases 1 to 163690)
 Rowen, L., Cameron, R.A. and Davidson, E.H.
 Direct Submission
 Submitted (23-AUG-2002) Multimegabase Sequencing Center, Institute
 for Systems Biology, 1441 N. 34th Street, Seattle, WA 98103, USA
 COMMENT
 Center: Multimegabase Sequencing Center
 Center code: ISBMS
 Web site: http://www.systemsbio.org
 Contact: leetowensystemsbio.org
 Drafting center: ISBMS
 Summary Statistics
 Sequencing vector: pUC18; 108752
 Chemistry: Dye-terminator Big Dye; 90% of reads
 Assembly program: Phrap; version 0.990399

NOTE: This is a 'working draft' sequence. It currently
 consists of 7 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces
 is believed to be correct as given, however the sizes
 of the gaps between them are based on estimates that have
 been provided by the submitter.
 This sequence will be replaced
 by the finished sequence as soon as it is available and
 the accession number will be preserved.
 1 41287: contig of 41287 bp in length
 41288 41387: gap of unknown length

41388 106251: contig of 64864 bp in length
 * 106252 106351: gap of unknown length
 * 106352 129007: contig of 22656 bp in length
 * 129008 129107: gap of unknown length
 * 129108 138703: contig of 9596 bp in length
 * 138704 138803: gap of unknown length
 * 138804 150335: contig of 11736 bp in length
 * 150336 160458: gap of unknown length
 * 160459 160558: gap of unknown length
 * 160559 163690: contig of 3132 bp in length.
 Location/Qualifiers
 source
 1. 163690
 /organism="Strongylocentrotus purpuratus"
 /db_xref="taxon:7668"
 /clone="sp58L24"
 /clone_lib="Caltech Strongylocentrotus purpuratus sperm
 genomic BAC library A"
 /note="Caltech Strongylocentrotus purpuratus sperm genomic
 BAC library A"
 BASE COUNT 51095 a 28957 c 32280 g 50747 t 611 others
 ORIGIN

Query Match 3.9%; Score 58.4; DB 2; Length 163690;
 Best Local Similarity 44.3%; Pred. No. 0.00077;
 Matches 239; Conservative 0; Mismatches 301; Indels 0; Gaps 0;
 QY 444 ACCAACAACCGAGCTCTCATGACACTATGATATTTACAGGAGATCACACCA 503
 DB 72297 AACATGCCACCAAGAAATTCCTGATTAAGACTATATTCGATGATGATG 72356
 QY 504 TCGTATATCCCGACGATGATGATGATGATGATGATGATGATGATGATGAT 563
 DB 72357 TCATATATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 72416
 QY 564 AGAACAACGACCTCTCTATCATCAGAGTCAGAGAAATTAACAGCATTTAGCCCA 623
 DB 72417 TCATAGCATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 72476
 QY 624 TTGTGACCTAAGATCTAATACACACCAACCAAGGCTCTTCATCATGTCACAGCA 683
 DB 72477 TCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 72536
 QY 684 CAGCATTTCAACAAGTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 743
 DB 72537 TCAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 72596
 QY 744 TAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 803
 DB 72597 GCATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 72656
 QY 804 CCTGACACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 863
 DB 72657 TCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 72716
 QY 864 TGAGTTGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 923
 DB 72717 TCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 72776
 QY 924 CAGATATTCGTCGATGATGATGATGATGATGATGATGATGATGATGATGATG 983
 DB 72777 TCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 72836

RESULT 13
 AC099341 51939 bp DNA linear PRI 22-FEB-2002
 LOCUS Homo sapiens BAC clone RP11-317C13 from 7, complete sequence.
 DEFINITION AC099341
 ACCESSION AC099341.4 GI:18464301
 VERSION AC099341.4
 KEYWORDS HTG; Homo sapiens.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
1 (bases 1 to 51939)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
2 (bases 1 to 51939)
Vanbrunt, A., Kozlowicz, A. and Boyer, F.
The sequence of Homo sapiens BAC clone RP11-317C13
Unpublished (2001)

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
3 (bases 1 to 51939)
Waterston, R.H.
Direct Submission
Submitted (09-NOV-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
4 (bases 1 to 51939)
Waterston, R.H.
Direct Submission
Submitted (01-FEB-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
5 (bases 1 to 51939)
Waterston, R.
Direct Submission
Submitted (22-FEB-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Feb 1, 2002 this sequence version replaced g1.17977469.

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.wustl.edu

Summary Statistics

Center project name: H_NH0317C13

COMMENT

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RP11-317C13 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Moon, P. Y., Zhao, B., Frengen, E., Tatenno, M., Catenease, J. J. and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org> VECTOR: pBac3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP13-620M21, 2000 bp overlap; the clone sequenced to the right is RP5-99414, 2000 bp overlap. Actual start of this clone is at base position 148410 of RP13-620M21.

There is a single plasmid subclone from 18209 to 18268. There is an unresolved simple sequence repeat from 3084 to 3812. Polymorphisms have been identified between AC104594, AC007129 and AC099341. Data from AC099649 was used to finish this clone, AC099341.

FEATURES

source
1..51939
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP11-317C13"
/clone_11b="RP11-11"
Location/Qualifiers
1..51939
/repeat_region
1..362
/repeat_region
489..542
/repeat_region
537..715
/repeat_region
724..1094
/repeat_region
2169..2478
/repeat_region
2172..2194
/repeat_region
2857..3072
/repeat_region
3095..3543
/repeat_region
3562..3807
/repeat_region
4215..4280
/repeat_region
4849..4894
/repeat_region
4895..5090
/repeat_region
5234..5369
/repeat_region
5364..5793
/repeat_region
5804..5900
/repeat_region
5906..6029
/repeat_region
6045..6159
/repeat_region
6161..6242
/repeat_region
6256..6437
/repeat_region
6442..6498
/repeat_region
6493..6588
/repeat_region
6754..7105
/repeat_region
7626..7746
/repeat_region
8074..8175
/repeat_region
10558..10645
/repeat_region
13385..13771
/repeat_region
15142..15777
/repeat_region
15939..15966
/repeat_region

/rpl_family="AT-rich"
15974. .16278
repeat_region /rpl_family="Alu"
17061. .17194
repeat_region /rpl_family="Mariner"
17279. .17479
repeat_region /rpl_family="(TG)n"
17513. .17641
repeat_region /rpl_family="(TG)n"
17653. .17778
repeat_region /rpl_family="(TG)n"
19254. .19432
repeat_region /rpl_family="MIR"
19464. .19518
repeat_region /rpl_family="MIR"
19496. .19803
repeat_region /rpl_family="(TTTA)n"
19916. .20118
repeat_region /rpl_family="Alu"
20484. .20627
repeat_region /rpl_family="Mariner"
21712. .21942
repeat_region /rpl_family="MIR"
21936. .21974
repeat_region /rpl_family="MIR"
21969. .22095
repeat_region /rpl_family="(CAT)n"
22833. .23064
repeat_region /rpl_family="L2"
22926. .22929
repeat_region /rpl_family="L1"
/note="match to EST R4562 (NID:9823776) y944c03.s1"
misc_feature
23065. .23371
repeat_region
23346. .23370
repeat_region /rpl_family="Alu"
23372. .23644
repeat_region /rpl_family="(CAAAA)n"
23751. .24052
repeat_region /rpl_family="L1"
23752. .23771
repeat_region /rpl_family="Alu"
23838. .24245
misc_feature /rpl_family="(TTG)n"
/note="similar to Homo sapiens EST BF933862 (NID:912351290)"
repeat_region
25280. .25317
misc_feature /rpl_family="(T)n"
25986. .26441

Query Match 3.88; Score 57.2; DB 9; Length 51939;
Best Local Similarity 44.38; Pred. No. 0.0014;
Matches 285; Conservative 0; Mismatches 353; Indels 6; Gaps 1;
QY 280 ATGCTTGGTACATGTCAGCAGGTTGATTCTATTGGCGAACTGTACATGCTCACCA 339
DB 3090 ATCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3149
QY 340 GAAAAACATGCGACCCCTGTCGACCTGTCGACCTGTCGACCTGTCGACCTGTCGACCTGTCG 399
DB 3150 GATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3209
QY 400 AATGAGATTGATCATATGCTTGGTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 459
DB 3210 GAT 3269
QY 460 CTTATGTCGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 519
DB 3270 CCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3329
QY 520 AATGAGATTGATCATATGCTTGGTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 579
DB 3330 GAT 3389
QY 580 CTACATGACGACTGAGGAAAGATAAACTGACCATTTATGCCCATTTGACAGTAAAGATCT 639

DB 3390 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3449
QY 640 AATGACACCAACCA-----AAGCCTTCATCATCATGATGCCGGAACCATTCATTC 693
DB 3450 GATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3509
QY 694 AATGAT 753
DB 3510 GAT 3569
QY 754 GATGACCAAAAGATCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 813
DB 3570 GAT 3629
QY 814 GGAT 873
DB 3630 GAT 3689
QY 874 GAT 917
DB 3690 GAT 3733

RESULT 14
AC104805
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AC104805 176480 bp DNA linear HMG 31-JAN-2002
Homo sapiens chromosome 4 clone RP11-371N1, WORKING DRAFT SEQUENCE,
3 unordered pieces.
AC104805
AC104805.2 GI:18450212
HMG: HMG5_PHASE1; HMG5_DRAFT; HMG5_ACTIVEFIN.
Homo sapiens.
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 176480)
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 176480)
Waterston, R.H.
Direct Submission
Submitted (21-DEC-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jan 31, 2002 this sequence version replaced gi:17975440.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
Project Information
Center project name: H_NH0571N01
----- Summary Statistics -----
Sequencing vector: M13; %
Sequencing vector: plasmid; %
Chemistry: Dye-primer ET; % of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 175267 bases at least Q40
Consensus quality: 175624 bases at least Q30
Consensus quality: 175694 bases at least Q20
Insert size: 19100; agarose-fp
Insert size: 176280; sum-of-ctrls
Quality coverage: 9.22 in Q20 bases; sum-of-ctrls
Quality coverage: 10.02 in Q20 bases; sum-of-ctrls
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OK nucleic - nucleic search, using \$w model

Run on: November 29, 2002, 06:34:04 ; Search time 268 Seconds
(without alignments)
12528.841 Million cell updates/sec

Title: US-09-868-760-6
Perfect score: 1491
Sequence: 1 gaggagagcagtgtaacga.....gctgttatagcagcgaca 1491

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

```
1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*
```

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1490.6	100.0	1491	AAAA7150	DNA encoding a set
2	1490.6	100.0	1491	AAAA7151	DNA encoding a set
3	56.8	3.8	1083	AA576745	DNA encoding novel
4	52.6	3.5	583	ABA50512	Human breast cell
5	52.6	3.5	583	ABA58463	Human foetal liver
6	52.6	3.5	583	ABA35452	Probe #1318 for g
7	52.6	3.5	583	AAK15834	Human brain expres
8	52.6	3.5	583	AAK42604	Human bone marrow
9	52.6	3.5	583	AAI23356	Probe #13289 for g

10	52.6	3.5	583	AA148680	Probe #17366 used
11	52.6	3.5	583	AA108992	Probe #8983 used t
12	52.6	3.5	583	AB516657	Human genome-deriv
13	52.6	3.5	1959	ABA45382	Human breast cell
14	52.6	3.5	1959	ABA55972	Human foetal liver
15	52.6	3.5	1959	ABA25346	Probe #4012 for ge
16	52.6	3.5	1959	AAK04090	Human brain expres
17	52.6	3.5	1959	AAK29571	Human bone marrow
18	52.6	3.5	1959	AA114146	Probe #4079 for ge
19	52.6	3.5	1959	AA135531	Probe #3985 used t
20	52.6	3.5	1959	AA103994	Human genome-deriv
21	52.6	3.5	1959	AB504127	Human foetal liver
22	49.2	3.3	327	ABA71406	Human brain expres
23	49.2	3.3	327	AAK19726	Human bone marrow
24	49.2	3.3	327	AAK45746	Probe #20358 used
25	49.2	3.3	327	AA151672	Human genome-deriv
26	49.2	3.3	327	AB520024	DNA encoding novel
27	47.8	3.2	1794	AA578187	Human foetal liver
28	46.8	3.1	8244	ABA58874	Human genome-deriv
29	46.2	3.1	456	AAK07034	DNA encoding novel
30	46.2	3.1	456	AAK32775	Human foetal liver
31	46.2	3.1	456	AAK32775	Human bone marrow
32	46.2	3.1	456	AAK32775	Probe #7275 used t
33	46.2	3.1	456	AB507574	Human genome-deriv
34	45	3.0	15935	AB507574	Probe #7275 used t
35	44.8	3.0	5340	AAK26289	Human genome-deriv
36	44.2	3.0	574	AB51944	Human foetal liver
37	44.2	3.0	574	AAK21762	Human foetal liver
38	44.2	3.0	574	AAK0231	Human brain expres
39	44.2	3.0	574	AAK25674	Human bone marrow
40	44.2	3.0	574	AA110301	Probe #234 for gen
41	44.2	3.0	574	AA131552	Probe #238 used to
42	44.2	3.0	574	AA100237	Probe #228 used to
43	44.2	3.0	574	AB500249	Human genome-deriv
44	44.2	3.0	669	ABA46726	Human breast cell
45	44.2	3.0	669	ABA64604	Human foetal liver

ALIGNMENTS

RESULT 1	AAAA7150 standard; DNA; 1491 BP.
ID	AAAA7150;
AC	AAAA7150;
DE	03-OCT-2000 (first entry)
DE	DNA encoding a serine protease inhibitor protein.
KW	Serine protease inhibitor; green-lipped mussel; anti-thrombin;
KW	divalent metal cation binding activity; dietary supplement;
KW	anticoagulant; ss.
OS	Perna canaliculus.
XX	WO200039165-A1.
XX	06-JUL-2000.
XX	23-DEC-1999; 99WO-N200227.
XX	23-DEC-1998; 98NZ-0333568.
XX	23-JUL-1999; 99NZ-0336906.
XX	(HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.
XX	Scotti PD, Dearing SC, Greenwood DR, Newcomb RD;
XX	WPI: 2000-452375/39.
XX	P-PSDB: AAY93750.

PT New Perna canaliculus serine protease inhibitor protein exhibiting
 PT anti-thrombin activity and divalent metal cation binding activity,
 PT useful as an anticoagulant agent and as a dietary supplement -
 XX
 PS Claim 10: Page 9-10; 44pp; English.
 CC The present sequence encodes a serine protease inhibitor
 CC protein. The protein is isolated from the green-lipped mussel
 CC (Perna canaliculus), and exhibits, inter alia, anti-thrombin activity
 CC and divalent metal cation binding activity. The serine protease
 CC inhibitor protein has a molecular weight of about 55 kilo Daltons.
 CC The protein, and its fragments, are useful in medicaments, in food,
 CC as dietary supplements or as bioremediation agents. In the dietary
 CC supplements, the protein is associated with or bound to at least one
 CC divalent cation (such as calcium, magnesium or zinc) of dietary
 CC significance. The proteins or their fragments are also useful as
 CC anticoagulant agents.
 CC
 XX
 SO Sequence 1491 BP: 428 A: 333 C: 350 G: 379 T: 1 other:

Query Match 100.0%; Score 1490.6; DB 21; Length 1491;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GAYGGGAGCAGTGTACGATGGGCGAGAACAAAGATGACACACACACACACACAT 60
 DB 1 GAYGGGAGCAGTGTACGATGGGCGAGAACAAAGATGACACACACACACACAT 60
 OY 61 GATCACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATGAT 120
 DB 61 GATCACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATGAT 120
 OY 121 GAACCAACCCCTGATATGCTAGCAGCCCTGACACACATGTCATGACATAGAGTTG 180
 DB 121 GAACCAACCCCTGATATGCTAGCAGCCCTGACACACATGTCATGACATAGAGTTG 180
 OY 181 TCACAGAAAGGTCATGAGCTGTTATATGAGAACTTCCTCTTGGGATTCACACAAAT 240
 DB 181 TCACAGAAAGGTCATGAGCTGTTATATGAGAACTTCCTCTTGGGATTCACACAAAT 240
 OY 241 GAAGACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATGATG 300
 DB 241 GAAGACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATGATG 300
 OY 301 GGTGTCATATCTATGCGCACTGTACATGCTCACCCAGAAAACATGCTGACCTGGT 360
 DB 301 GGTGTCATATCTATGCGCACTGTACATGCTCACCCAGAAAACATGCTGACCTGGT 360
 OY 361 GACCTCGGTGACCTGCTGACGATGATGATGATGATGATGATGATGATGATGATG 420
 DB 361 GACCTCGGTGACCTGCTGACGATGATGATGATGATGATGATGATGATGATGATG 420
 OY 421 TGGTTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480
 DB 421 TGGTTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480
 OY 481 ATTTTACAAAGGATGACACACATGCTGATACCCAGCAGATAGAAATCGCTGTGT 540
 DB 481 ATTTTACAAAGGATGACACACATGCTGATACCCAGCAGATAGAAATCGCTGTGT 540
 OY 541 ATTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
 DB 541 ATTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
 OY 601 GATTAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
 DB 601 GATTAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
 OY 661 CTTCATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
 DB 661 CTTCATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
 OY 721 GTGTCTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780

DB 721 GTGTCTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780
 OY 781 GTACAGATCTACGCCCAACGGGTGACCTGACAGTGTGATGATMACCTGGTCTAAATAT 840
 DB 781 GTACAGATCTACGCCCAACGGGTGACCTGACAGTGTGATGATMACCTGGTCTAAATAT 840
 OY 841 GATCCTCATGAAAGATATACACAGTGTGATGATGATGATGATGATGATGATGATGAT 900
 DB 841 GATCCTCATGAAAGATATACACAGTGTGATGATGATGATGATGATGATGATGATGAT 900
 OY 901 CATGGCGTGTCAATGAAAGCAGATATCTCGATCAATATCTTGGTGATGACAGT 960
 DB 901 CATGGCGTGTCAATGAAAGCAGATATCTCGATCAATATCTTGGTGATGACAGT 960
 OY 961 GTCCTGCGACGTTCTATTTGCCATTACCAAGACCATCTTCATTAAGTCCAAATTT 1020
 DB 961 GTCCTGCGACGTTCTATTTGCCATTACCAAGACCATCTTCATTAAGTCCAAATTT 1020
 OY 1021 GCCTGTTGTCTATGACAGTGTGACAGAGCCATCCAGAAATGTTCTACAGACTAATGT 1080
 DB 1021 GCCTGTTGTCTATGACAGTGTGACAGAGCCATCCAGAAATGTTCTACAGACTAATGT 1080
 OY 1081 GTGTGACGCTAATACAGAAATCTACTGTTTACATCACCATGCTCTGTCTATTAACA 1140
 DB 1081 GTGTGACGCTAATACAGAAATCTACTGTTTACATCACCATGCTCTGTCTATTAACA 1140
 OY 1141 TTGGAACAGACCCCTGAGAGATCAACACATATGACGGCTGATCTCAAGATTTTACGTT 1200
 DB 1141 TTGGAACAGACCCCTGAGAGATCAACACATATGACGGCTGATCTCAAGATTTTACGTT 1200
 OY 1201 AGTGAAGACTTGTCAATCATATGATGATGATGATGATGATGATGATGATGATGAT 1260
 DB 1201 AGTGAAGACTTGTCAATCATATGATGATGATGATGATGATGATGATGATGATGAT 1260
 OY 1261 CATGGCTGTCACTCTTATAGGAGATGTACATGATGATGATGATGATGATGATGATG 1320
 DB 1261 CATGGCTGTCACTCTTATAGGAGATGTACATGATGATGATGATGATGATGATGATG 1320
 OY 1321 AGACCTGGTGAACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
 DB 1321 AGACCTGGTGAACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
 OY 1381 ACCTTTGATCATCTTAATGTTGAAGATCTTAACGACAGTTCCTTGTATATATGACAGGC 1440
 DB 1381 ACCTTTGATCATCTTAATGTTGAAGATCTTAACGACAGTTCCTTGTATATATGACAGGC 1440
 OY 1441 GGACATGAGGTGAGAGTGAAGAGGTTGCTTGTCTGTATATGACAGCGGCA 1491
 DB 1441 GGACATGAGGTGAGAGTGAAGAGGTTGCTTGTCTGTATATGACAGCGGCA 1491

RESULT 2
 AAA47151
 ID AAA47151 standard; DNA; 1611 BP.
 XX
 AC AAA47151:
 DT 03-OCT-2000 (first entry)
 DE
 XX
 DE DNA encoding a serine protease inhibitor protein.
 XX
 KW Serine protease inhibitor; green-lipped mussel; anti-thrombin;
 KW divalent metal cation binding activity; dietary supplement;
 KW anticoagulant; ss.
 XX
 OS Perna canaliculus.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1494
 FT /product= "serine protease inhibitor"
 FT polyA_site 1557..1563

ET /*tag= b
XX MO200039165-A1.
XX 06-JUL-2000.
XX 23-DEC-1999; 99MO-NZ00227.
XX 23-DEC-1998; 98NZ-0333568.
XX 23-JUL-1999; 99NZ-0336906.
XX (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.
XX Scotti PD, Dearing SC, Greenwood DR, Newcomb RD;
XX WPI: 2000-452375/39.
XX P-PSDB: AAY93750.
XX New Perna canaliculus serine protease inhibitor protein exhibiting
XX anti-chrombin activity and divalent metal cation binding activity,
XX useful as an anticoagulant agent and as a dietary supplement -
XX
XX Claim 11; Page 10-11; 44pp; English.
XX
XX The present sequence encodes a serine protease inhibitor
XX protein. The protein is isolated from the green-lipped mussel
XX (Perna canaliculus), and exhibits, inter alia, anti-thrombin activity
XX and divalent metal cation binding activity. The serine protease
XX inhibitor protein has a molecular weight of about 55 kilo Daltons.
XX The protein and its fragments are useful in medicaments, in food,
XX as dietary supplements or as bioremediation agents. In the dietary
XX supplements, the protein is associated with or bound to at least one
XX divalent cation (such as calcium, magnesium or zinc) of dietary
XX significance. The proteins or their fragments are also useful as
XX anticoagulant agents.
XX
XX Sequence 1611 BP; 499 A; 348 C; 360 G; 402 T; 2 other:
XX
XX Query Match 100.0%; Score 1490.6; DB 21; Length 1611;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 1491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GAYGGGAGCAGTGTACGATGGGCGAGAACAAAGATGACACCATGAGACACACGAT 60
DB 1 GAYGGGAGCAGTGTACGATGGGCGAGAACAAAGATGACACCATGAGACACACGAT 60
OY 61 GATGACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATGATG 120
DB 61 GATGACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATGATG 120
OY 121 GATGACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATGATG 180
DB 121 GATGACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATGATG 180
OY 181 TCACAGAAAGGCTGATGAGCTTTATCTAGAACTTCATCTTCGCGATTAACAGAA 240
DB 181 TCACAGAAAGGCTGATGAGCTTTATCTAGAACTTCATCTTCGCGATTAACAGAA 240
OY 241 GAGAGCCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATGATG 300
DB 241 GAGAGCCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATGATG 300
OY 301 GGTGTGATTTATGCGCACTGTACAAATGTCACCCGAGAAAACATGCGACCTGCT 360
DB 301 GGTGTGATTTATGCGCACTGTACAAATGTCACCCGAGAAAACATGCGACCTGCT 360
OY 361 GACCTCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
DB 361 GACCTCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
OY 421 TGGTGGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
DB 421 TGGTGGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480

OY 481 ATTTACANGGAGTCACACCGATGCTGATACCCGACGAGTGAATGCCGCTGTGCT 540
DB 481 ATTTACANGGAGTCACACCGATGCTGATACCCGACGAGTGAATGCCGCTGTGCT 540
OY 541 ATTGTCATGAGAAAAGCTGCGCCGAGAAAACAGACGCTGCTTACATCAGCATAGAGNA 600
DB 541 ATTGTCATGAGAAAAGCTGCGCCGAGAAAACAGACGCTGCTTACATCAGCATAGAGNA 600
OY 601 GATGAAAAGTGAACATTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
DB 601 GATGAAAAGTGAACATTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
OY 661 CTTCATCATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
DB 661 CTTCATCATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
OY 721 GTGTCCTACCATTTAGAGGATTTATGATGATGATGATGATGATGATGATGATGATGAT 780
DB 721 GTGTCCTACCATTTAGAGGATTTATGATGATGATGATGATGATGATGATGATGATGAT 780
OY 781 GTACAGATCTAGCGCCAGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
DB 781 GTACAGATCTAGCGCCAGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
OY 841 GATGTCATGAGAAATTTACACGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
DB 841 GATGTCATGAGAAATTTACACGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
OY 901 CATGGGCTGTCATGAAAGCAGAGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
DB 901 CATGGGCTGTCATGAAAGCAGAGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
OY 961 GTGCTGGAGCTTCTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
DB 961 GTGCTGGAGCTTCTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
OY 1021 GCTGTTGCTGCTATGAGAGCTGAGAGCCATCCAGAAATTTGCTGCTGCTGCTGCTGCT 1080
DB 1021 GCTGTTGCTGCTATGAGAGCTGAGAGCCATCCAGAAATTTGCTGCTGCTGCTGCTGCT 1080
OY 1081 GTTGTGAGACCTATATACAAATCTAGTTTATGATGATGATGATGATGATGATGATGATGAT 1140
DB 1081 GTTGTGAGACCTATATACAAATCTAGTTTATGATGATGATGATGATGATGATGATGATGAT 1140
OY 1141 TTGGAACAGACCCCTGAGAGATCAACATATGAGCGGCTGCTGCTGCTGCTGCTGCTGCT 1200
DB 1141 TTGGAACAGACCCCTGAGAGATCAACATATGAGCGGCTGCTGCTGCTGCTGCTGCTGCT 1200
OY 1201 AGTGAAGACTTGTACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260
DB 1201 AGTGAAGACTTGTACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260
OY 1261 CATGGCTGCTGCTTATGAGAGATTTGATGATGATGATGATGATGATGATGATGATGATG 1320
DB 1261 CATGGCTGCTGCTTATGAGAGATTTGATGATGATGATGATGATGATGATGATGATGATG 1320
OY 1321 AGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
DB 1321 AGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
OY 1381 ACCCTTGTATCATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1440
DB 1381 ACCCTTGTATCATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1440
OY 1441 GGACATGAGTGTGAGAGTGTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1491
DB 1441 GGACATGAGTGTGAGAGTGTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1491
RESULT 3
AAS76745
ID AAS76745 standard; cDNA: 1083 BP.

XX AAS76745;
 XX
 XX 13-FEB-2002 (first entry)
 XX
 XX DNA encoding novel human diagnostic protein #12549.
 DE
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001: 2001WO-US08631.
 XX
 PR 31-MAR-2000: 2000US-0540217.
 PR 23-AUG-2000: 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Dermanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR P-PSDB: ABG12558.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 1; SEQ ID NO 12549; 103pp; English.
 XX
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and a
 CC food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94364 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).
 XX
 XX Sequence 1083 BP; 304 A; 109 C; 375 G; 295 T; 0 other;

[illegible]

Db	379	GATGTCGACAAATGATATGATGATGGTGATTAATGATGATTAATGATTAATGATGATGAT	438
Oy	181	TCACAGAAAGGTCATGAGACGCTTATCTGAACCTCATTCTTGCGGATTCACACAACT	240
Db	439	GATGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	498
Oy	241	GAAGACCATGACGACACCACTCATGAGACTTCTGACACATGCTTGTCACATGTCAGCA	300
Db	499	GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	558
Oy	301	GGTTGTGATTCATTATTGGCGAACTGTACATGCTCACCCAGAAAAACATGCTGACCC	360
Db	559	AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	618
Oy	361	GACCTGCGTACGACCTGCTGTTGACGATGATAGGCGCGCTGTTAATGAATGATCATCAT	420
Db	619	GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	678
Oy	421	TG 422	
Db	679	GG 680	
RESULT 4			
ABAS0512			
ID	ABAS0512	standard; DNA; 583 BP.	
XX	ABAS0512:		
XX	01-FEB-2002	(first entry)	
XX	Human breast cell single exon nucleic acid probe #9207.		
XX	Human; microarray; single exon probe; gene expression; breast;		
XX	disease; cancer; ss.		
XX	Homo sapiens.		
XX	MO200157271-A2.		
XX	09-AUG-2001.		
XX	30-JAN-2001; 2001WO-US00662.		
XX	04-FEB-2000; 2000US-0180312.		
XX	26-MAY-2000; 2000US-0207456.		
XX	30-JUN-2000; 2000US-0608408.		
XX	03-AUG-2000; 2000US-0632366.		
XX	21-SEP-2000; 2000US-0234687.		
XX	27-SEP-2000; 2000US-0236359.		
XX	04-OCT-2000; 2000GB-0024263.		
XX	(MOLE-) MOLECULAR DYNAMICS INC.		
XX	Penn SG, Hanzel DK, Chen W, Rank DR;		
XX	WPI; 2001-496933/54.		
XX	New spatially-addressable set of single exon nucleic acid probes,		
XX	useful for measuring gene expression in sample derived from human		
XX	breast, comprises number of single exon nucleic acid probes		
XX	Claim 4; SEQ ID NO 9207; 327bp + sequence listing; English.		
XX	The invention relates to a spatially-addressable set of single exon		
XX	nucleic acid probes for measuring gene expression in a sample derived		
XX	from human breast and BT 474 cells. The method involves contacting		
XX	the probes with a collection of detectably labeled nucleic acids		
XX	derived from mRNA of human breast, and then measuring the label		
XX	bound to each probe of the microarray. The probes are useful for		
XX	verifying the expression of regions of genomic DNA predicted to		
XX	encode proteins. They are useful for gene discovery, and for		
XX	determining predisposition and/or prognosing breast disease. Gene		

XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00666.
PF
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI: 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
XX Claim 4: SEQ ID No 13918; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging
CC monitoring and prognosis diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp://wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 583 BP; 172 A; 38 C; 212 G; 161 T; 0 other;

Query Match 3.5%; Score 52.6; DB 22; Length 583;
Best Local Similarity 47.0%; Pred. No. 7.8e-05;
Matches 194; Conservative 1; Mismatches 215; Indels 3; Gaps 1;
QY 1 GAYGGGAGCAGTGTACGATGGGCGACACAAAGATGACCCACCATGACACCCAGCAT 60
DB 41 GATGGGAGT 100
QY 61 GATCACCATGACGACATGT 120
DB 101 GATGT 160
QY 121 GAACCAACCCCTCATTTAGCTAGACGCTTACACCATGTCATGGCAGCATAGAGTGTG 180
DB 161 GGGGAAGATGT 220
QY 181 TCACGAAAGGCTGATGAGCTTTTATCTAGACTTCTGTCATCTGTGTGTGTGTGTGT 240
DB 221 ACTGATGT 280
QY 241 GAAGCAGT 300
DB 281 GATGT 337
QY 301 GGT 360
DB 338 GATGT 397
QY 361 GACCTCGGT 413
DB 398 ATGT 450

RESULT 7

AAK16834
ID AAK16834 standard; DNA; 583 BP.
XX
XX AAK16834;
AC
XX
XX 05-NOV-2001 (first entry)
DT
XX
XX Human brain expressed single exon probe SEQ ID NO: 16825.
DE
XX
XX Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157275-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI: 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
XX
XX Example 4: SEQ ID NO: 16825; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 583 BP; 172 A; 38 C; 212 G; 161 T; 0 other;

Query Match 3.5%; Score 52.6; DB 22; Length 583;
Best Local Similarity 47.0%; Pred. No. 7.8e-05;
Matches 194; Conservative 1; Mismatches 215; Indels 3; Gaps 1;
QY 1 GAYGGGAGCAGT 60
DB 41 GATGGGAGT 100
QY 61 GATCACCATGACGACATGT 120
DB 101 GATGT 160
QY 121 GAACCAACCCCTCATTTAGCTAGACGCTTACACCATGTCATGGCAGCATAGAGTGTG 180
DB 161 GGGGAAGATGT 220
QY 181 TCACGAAAGGCTGATGAGCTTTTATCTAGACTTCTGTCATCTGTGTGTGTGTGTGT 240
DB 221 ACTGATGT 280
QY 241 GAAGCAGT 300
DB 281 GATGT 337

```

Oy 301 GGTGTGATTCATTATGCGACATGATGTCACCCAGAAAAACATGCTGACCTGCG 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 338 GATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 397
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 361 GACCTCGTACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 398 AATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 450
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
AAK42604
ID AAK42604 standard; DNA; 583 BP.
XX
XX AAK42604;
AC
XX 06-NOV-2001 (first entry)
DT
XX
XX Human bone marrow expressed single exon probe SEQ ID NO: 17161.
DE
XX Human: bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200157276-A2.
PN
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00668.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-488900/53.
PT
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
XX Example 4: SEQ ID NO: 17161; 658bp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is one of
XX the probes of the invention.
XX
XX Sequence 583 BP; 172 A; 38 C; 212 G; 161 T; 0 other;
XX

Query Match 3.5%; Score 52.6; DB 22; Length 583;
Best Local Similarity 47.0%; Pred. No. 7.8e-05;
Matches 194; Conservative 1; Mismatches 215; Indels 3; Gaps 1;

```

```

Db 161 GGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 220
Oy 181 TCACAGAGGGGTCATGAGGCTGTTTATCTTGAACCTTCATCTTGTGATGATGATGATGAT 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 221 ACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 280
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 241 GAAGACCATGACGACACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 281 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 337
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 301 GGTGTGATTCATTATGCGAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 338 GATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 397
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 361 GACCTCGTACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 398 AATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 450
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
AAI23356
ID AAI23356 standard; DNA; 583 BP.
XX
XX AAI23356;
AC
XX
XX 12-OCT-2001 (first entry)
DT
XX
XX Probe #13289 for gene expression analysis in human cervical cell sample.
DE
XX
XX Probe: human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200157278-A2.
PN
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00670.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-488901/53.
PT
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 25: SEQ ID NO 13289; 487bp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENPs). The present sequence is one such probe. The SENPs are derived
XX from human HeLa cells. The SENPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging
XX of diseases of the cervix, notably cervical cancer.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp://ipb.int/published_pcl_sequences.
XX
XX Sequence 583 BP; 172 A; 38 C; 212 G; 161 T; 0 other;
XX

Query Match 3.5%; Score 52.6; DB 22; Length 583;

```

Best Local Similarity 47.0%; Pred. No. 7.8e-05;
Matches 194; Conservative 1; Mismatches 215; Indels 3; Gaps 1;

```
OY 1 GAYGGGAGCAGTGTAAACGATGGGAGCAACAAAGATGACACCATGACACACACAGAT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 41 GATGGGATGATGGTGGATGGTGGATGGATGACAGTGGATGGATGGATGGATGGG 100
OY 61 GATCACCATGACGACACCATGATGATGATGATGATGATGATGATGATGATGATG 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 101 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 160
OY 121 GACCAACACCTCATATGGCTAGACGACCTTACACCATGTCATGACATGAGATTG 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 161 GGGGAAGATGATGACAGATGATGATGATGATGATGATGATGATGATGATGATG 220
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 181 TCACAGAAAGGTCATGAGCTGTTTATCTAGAACTTCATCTTGTGATTCACACA 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 221 AGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 280
OY 241 GAAACCATGACGACACCATGATGATGATGATGATGATGATGATGATGATGATG 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 281 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 337
OY 301 GGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 338 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 397
OY 361 GACCTCGGTGACCTGTTGATGATGATGATGATGATGATGATGATGATGATG 413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 398 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 450
```

RESULT 10

AA148680 standard; DNA: 583 BP.

AC AA148680;

DT 17-OCT-2001 (first entry)

DE Probe #17366 used to measure gene expression in human placenta sample.

KW Probe; microarray; human; placenta; antenatal diagnosis;

KW genetic disorder; ss.

OS Homo sapiens.

PN WO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

PI WPI; 2001-488897/53.

PT Human genome-derived single exon nucleic acid probes useful for

PS Claim 25; SEQ ID NO 17366; 654bp; English.

CC The present invention relates to single exon nucleic acid probes (SENP).

CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.

Sequence 583 BP; 172 A; 38 C; 212 G; 161 T; 0 other:

Query Match 3.5%; Score 52.6; DB 22; Length 583;

Best Local Similarity 47.0%; Pred. No. 7.8e-05;
Matches 194; Conservative 1; Mismatches 215; Indels 3; Gaps 1;

```
OY 1 GAYGGGAGCAGTGTAAACGATGGGAGCAACAAAGATGACACCATGACACACAGAT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 41 GATGGGATGATGGTGGATGGTGGATGGATGACAGTGGATGGATGGATGGATGGG 100
OY 61 GATCACCATGACGACACCATGATGATGATGATGATGATGATGATGATGATGATG 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 101 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 160
OY 121 GACCAACACCTCATATGGCTAGACGACCTTACACCATGTCATGACATGAGATTG 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 161 GGGGAAGATGATGACAGATGATGATGATGATGATGATGATGATGATGATGATG 220
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 181 TCACAGAAAGGTCATGAGCTGTTTATCTAGAACTTCATCTTGTGATTCACACA 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 221 AGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 280
OY 241 GAAACCATGACGACACCATGATGATGATGATGATGATGATGATGATGATGATG 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 281 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 337
OY 301 GGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 338 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 397
OY 361 GACCTCGGTGACCTGTTGATGATGATGATGATGATGATGATGATGATGATG 413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 398 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 450
```

RESULT 11

AA108992 standard; DNA: 583 BP.

AC AA108992;

DT 09-OCT-2001 (first entry)

DE Probe #8983 used to measure gene expression in human breast sample.

KW Probe; human; breast disease; breast cancer; development disorder; ss;

KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.

OS Homo sapiens.

PN WO200157270-A2.

PD 09-AUG-2001.

PF 29-JAN-2001; 2001WO-US00661.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

PI WPI; 2001-488897/53.

PT Human genome-derived single exon nucleic acid probes useful for

PS Claim 25; SEQ ID NO 17366; 654bp; English.

CC The present invention relates to single exon nucleic acid probes (SENP).

Mon Dec 2 12:19:29 2002

us-09-868-760-6.rng

Page 12

[illegible]

Search completed: November 29, 2002, 07:40:20
Job time : 289 secs

•

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2002, 06:39:33 ; Search time 50 Seconds
(without alignments)
11484.142 Million cell updates/sec

Title: US-09-868-760-6

Perfect score: 1491
Sequence: 1 gayggggagcagtgtaacga.....gctgtgtatagacgggca 1491

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 341543 segs, 192557720 residues

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PC7_NEM_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEM_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_NEM_PUB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEM_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PC7US_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEM_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEM_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEM_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEM_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	52.6	3.5	583	US-09-864-761-20772	Sequence 20772, A
2	52.6	3.5	1959	US-09-864-761-4012	Sequence 4012, A
3	49.2	3.3	327	US-09-864-761-28059	Sequence 28059, A
4	46.2	3.1	456	US-09-864-761-11468	Sequence 11468, A
5	44.2	3.0	574	US-09-864-761-228	Sequence 228, App
6	44.2	3.0	659	US-09-864-761-17051	Sequence 17051, A
7	43.4	2.9	470	US-09-864-761-3121	Sequence 3121, App
8	43.4	2.9	522	US-09-864-761-19900	Sequence 19900, A
9	43.2	2.9	537	US-09-864-761-8330	Sequence 8330, App
10	43.2	2.9	1075	US-09-864-761-19241	Sequence 19241, A
11	43.2	2.9	1403	US-09-864-761-2513	Sequence 2513, App
12	42.6	2.9	1959	US-09-864-761-4012	Sequence 4012, App
13	42.4	2.8	703	US-09-910-943-302	Sequence 302, App
14	42.4	2.8	766	US-09-864-761-19608	Sequence 19608, A
15	42.4	2.8	1944	US-09-864-761-2825	Sequence 2825, App
16	41.4	2.8	276	US-09-864-761-25120	Sequence 25120, A
17	41	2.7	240	US-09-864-761-23948	Sequence 23948, A
18	41	2.7	554	US-09-864-761-7217	Sequence 7217, App
19	40.6	2.7	390	US-09-790-399-7	Sequence 7, Appl1

20	40.4	2.7	672	US-09-938-842A-2202	Sequence 2202, App
21	40.2	2.7	366	US-09-974-300-6355	Sequence 6355, App
22	39.8	2.7	2403	US-09-815-242-9287	Sequence 9287, App
23	39.6	2.7	2069	US-09-842-552-17	Sequence 17, Appl1
24	38.8	2.6	439	US-09-864-761-20117	Sequence 20117, A
25	38.8	2.6	441	US-09-864-761-3403	Sequence 3403, App
26	38.8	2.6	444	US-09-864-761-17518	Sequence 27518, A
27	38.8	2.6	478	US-09-728-446-1415	Sequence 1415, App
28	38.8	2.6	491	US-09-864-761-10879	Sequence 10879, A
29	38.8	2.6	517	US-09-920-300A-313	Sequence 313, App
30	38.8	2.6	517	US-10-033-528-313	Sequence 313, App
31	38.8	2.6	20029	US-09-764-847-17053	Sequence 17053, A
32	38.6	2.6	331	US-09-864-761-17053	Sequence 17053, A
33	38.6	2.6	446	US-09-864-761-20699	Sequence 20699, A
34	38.6	2.6	465	US-09-864-761-230	Sequence 230, App
35	38.4	2.6	2581	US-09-938-842A-592	Sequence 592, App
36	38.2	2.6	423	US-09-864-761-8619	Sequence 8619, App
37	38.2	2.6	565	US-09-864-761-20772	Sequence 20772, A
38	38.2	2.6	583	US-09-864-761-18475	Sequence 18475, A
39	37.8	2.5	131	US-09-864-761-18475	Sequence 1, Appl1
40	37.8	2.5	258	US-10-028-247-1	Sequence 1720, App
41	37.8	2.5	491	US-09-864-761-1720	Sequence 2326, A
42	37.4	2.5	420	US-09-864-761-2326	Sequence 6534, App
43	37.4	2.5	451	US-09-864-761-6554	Sequence 4184, App
44	37.4	2.5	473	US-09-864-761-4184	Sequence 5, Appl1
45	37.4	2.5	1411	US-09-191-724-5	

ALIGNMENTS

RESULT 1
US-09-864-761-20772
Sequence 20772, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecolica X-1
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30

```
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 20772
;; LENGTH: 583
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL022334.1
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.1
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
;; OTHER INFORMATION: EXPRESSED IN HELI00, SIGNAL = 2.3
;; OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 2.7
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
;; OTHER INFORMATION: EST_HUMAN HIT: AW92395.1, EVALUO 9.50e-02
US-09-864-761-20772
```

```
Query Match 3.5%; Score 52.6; DB 10; Length 583;
Best Local Similarity 47.0%; Pred. No. 7.3e-06;
Matches 194; Conservative 1; Mismatches 215; Indels 3; Gaps 1;
```

```
QY 1 GAYGGGAGCAGTGTAAAGTGGGACCAACAAAGATGACACCATGACGACGACACCATGAT 60
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 41 GATGGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 100
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 GATCACCATGACGACGACGATGATGATGATGATGATGATGATGATGATGATGATGATG 120
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 101 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 160
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 GAACCAACCCCTCATATGCTAGCAGCCTTACACCATGCTGATGATGATGATGATGATG 180
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 161 GCGGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 220
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 181 TCACGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 221 ACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 280
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 241 GAAGACCATGACGACGACGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 281 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 337
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 301 GGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 338 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 397
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 361 GACCTCGGTCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 413
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 398 AATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 450
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 2

```
US-09-864-761-4012
;; Sequence 4012, Application US/09864761
;; Patent No. US20020048763A1
;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharon G.
;; APPLICANT: Rank, David R.
;; APPLICANT: Hanzel, David K.
;; APPLICANT: Chen, Wensheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
```

```
;; FILE REFERENCE: Aecm1ca-X-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263,6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 4012
;; LENGTH: 1959
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP NO AL022334.1
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.1
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
;; OTHER INFORMATION: EXPRESSED IN HELI00, SIGNAL = 2.3
;; OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 2.7
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
US-09-864-761-4012
```

```
Query Match 3.5%; Score 52.6; DB 10; Length 1959;
Best Local Similarity 47.0%; Pred. No. 1.6e-05;
Matches 194; Conservative 1; Mismatches 215; Indels 3; Gaps 1;
```

```
QY 1 GAYGGGAGCAGTGTAAAGTGGGACCAACAAAGATGACACCATGACGACGACACCATGAT 60
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 546 GATGGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 605
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 GATCACCATGACGACGACGATGATGATGATGATGATGATGATGATGATGATGATGATG 120
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 606 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 665
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 GAACCAACCCCTCATATGCTAGCAGCCTTACACCATGCTGATGATGATGATGATGATG 180
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 666 GGGGAATATATACAGTATGCTGCTGATGATGACATGACATGATGCGGATGATGAC 725
Oy 181 TCACAGAGGCTCATGAGCTGTTTATCTAGAACTTCCTTGTGATTCACACAGT 240
Db 726 AGTATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 785
Oy 241 GAAGACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
Db 786 GATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 842
Oy 301 GGTGTGATTCATGCTGACATGCTGACATGCTGACATGCTGACATGCTGACATGCTGAC 360
Db 843 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 902
Oy 361 GACCTGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 413
Db 903 AATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 955

RESULT 3
US-09-864-761-28059/c
Sequence 28059, Application US/09864761
Patent No. US2002048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 28059
LENGTH: 327

TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005668.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.55
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
US-09-864-761-28059

Query Match 3.3% Score 49.2; DB 10; Length 327;
Best Local Similarity 47.7%; Pred No 5.3e-05;
Matches 144; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

Oy 744 TATGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 803
Db 303 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 863
Oy 804 CCGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 863
Db 243 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 884
Oy 864 TGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 923
Db 183 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 983
Oy 924 CAGATATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 983
Db 123 CAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 983
Oy 984 TCACCAAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1043
Db 63 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1043
Oy 1044 AC 1045
Db 3 GC 2

RESULT 4
US-09-864-761-11468/c
Sequence 11468, Application US/09864761
Patent No. US2002048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
```



```

1 PRIOR FILING DATE: 2000-09-21
2 PRIOR APPLICATION NUMBER: US 09/608,408
3 PRIOR FILING DATE: 2000-06-30
4 PRIOR FILING DATE: 2000-01-29
5 PRIOR FILING DATE: 2000-01-29
6 NUMBER OF SEQ ID NOS: 49117
7 SOFTWARE: Anomax Sequence Listing Engine vers. 1.1.1
8 SEQ ID NO: 3121
9 LENGTH: 470

```

```

? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? OTHER INFORMATION: MAP TO AL049749.1
? OTHER INFORMATION: EXPRESSED IN BT4749, SIGNAL = 5.8
? OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.2
? OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.8
? OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
? OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.7
? OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.7
? OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.3
? OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.3
? OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.9
? OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 4.4
? OS-09-864-761-3121

```

Query Match	2.98;	Score 43.4;	DB 10;	Length 470;
Best Local Similarity	55.7%;	Pred. No. 0.0037;		
Matches 83; Conservative	0;	Mismatches 66;	Indels 0;	Gaps 0

Qy	17	ACATGGGCGAACAAGTGTGCATGAGACGACGACGATGATACCATGAGAC	76
Db	160	ACCATCATCAACACACCATCATCATCATACACACACACCATCACACACAC	10
Qy	77	ATGATGAATGAATGAAACAACAGCACTATGCCAGTGTAAATGGAACCAACCTCATA	13
Db	100	ACCTTCATGATCACTGCACCTGCATCTCATGCTTATTAATACACCGCATACCATCATTC	41
Qy	137	TGGCTAGACGCGCTTCACCAACATGTCAT	165
Db	40	ACTACCAACCATCATCAACATCACTACCAT	12

```

RESULT 8
US-09-864-761-19900/c
; Sequence 19900, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864, 761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669

```

```

PRIORITY FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,667
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
SEQ ID NO 19900
LENGTH: 522
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL049749.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.8
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.8
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.7
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.3
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN HL100, SIGNAL = 4.4
OTHER INFORMATION: EST_HUMAN HIT: AA776664.1, EVALUATE 4.90e-01
US-09-864-761-19900

Query Match      2.9%; Score 43.4; DB 10; Length 522:
Best Local Similarity 55.7%; Pred. No. 0.004; Indels 0; Gaps 0;
Matches 83; Conservative 0; Mismatches 66;

QY 17 ACGATGGCAGCAACAAGAATGATGCACCATGTACGACGCCACGATGATCACCATGACGACC 76
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 421 ACGATGATGATGCACCCACCATGTCATGTCACGACGCCACCATGTCACCATGTCACCATGTCAC 362
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 77 ATGATGATGATGATGAACAAATGCACTACTATGTCGCCAGTGTGAATGGAACCAAAACCTTCATA 136
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 361 ACCCTCATGATGATGTCAGTGCCACCTGTCATGTCATGTCATTAATTATTCACGCCGTCATCCATCATC 302
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 137 TGGCTAGCAGCCTTCACCCACCATGTGCAT 165
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 301 ACTACCCACCATCATCATCCATCATCACAAT 273
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
US-09-864-761-8390
Sequence 8390, Application us/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon R.
APPLICANT: Bank, David G.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aomicia X-1
CURRENT APPLICATION NUMBER: US 09/864,761
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
```



```

; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 19608
; LENGTH: 766
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: MAP TO AL008720.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL - 1.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL - 0.89
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL - 1.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL - 1.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL - 1.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL - 1.2
; OTHER INFORMATION: SWISSPROT HIT: P17164, EVALU 4.10e+00
; OTHER INFORMATION: EST_HUMAN HIT: AW844901.1, EVALU 2.90e+00
US-09-864-761-19608

```

```

Query Match      2.8%; Score 42.4; DB 10; Length 766;
Best Local Similarity 46.9%; Pred. No. 0.01;
Matches 199; Conservative 0; Mismatches 221; Indels 4; Gaps 2;

```

```

OY 19 GATGGGCAAGCAAGATGACACCATGACGACCAACCATGATGACATGACGACCAT 78
DB 621 GAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 111
OY 79 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 138
DB 561 GAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 502
OY 139 -GGTACAGCCTTCACACCATGATGATGATGATGATGATGATGATGATGATG 196
DB 501 GTGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 442
OY 197 GAGCTGTTATCTGAACTTCACTTGTGGATTCAACACAGAGAGACCATGACGAC 256
DB 441 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 382
OY 257 ACCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 314
DB 381 ACAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 322
OY 315 TGGGCAACTGTACATGCTCCACCAAGAAACATCTGATGATGATGATGATGATG 374
DB 321 TGACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 262
OY 375 GATTGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 434
DB 261 TGTGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 202
OY 435 TGGT 438
DB 201 TGGT 198

```

```

RESULT 15
US-09-864-761-2825/C
; Sequence 2825, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23

```

```

; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 1944
; LENGTH: 1944
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: MAP TO AL008720.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL - 1.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL - 0.89
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL - 1.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL - 1.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL - 1.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL - 1.2
US-09-864-761-2825

```

```

Query Match      2.8%; Score 42.4; DB 10; Length 1944;
Best Local Similarity 46.9%; Pred. No. 0.018;
Matches 199; Conservative 0; Mismatches 221; Indels 4; Gaps 2;

```

```

OY 19 GATGGGCAAGCAAGATGACACCATGACGACCAACCATGATGACATGACGACCAT 78
DB 858 GAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 799
OY 79 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 138
DB 798 GAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 739
OY 139 -GCTAGCAGCCTTCACACCATGATGATGATGATGATGATGATGATGATGATG 196
DB 738 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 679
OY 197 GAGCTGTTATCTGAACTTCACTTGTGGATTCAACACAGAGAGACCATGACGAC 256
DB 678 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 619
OY 257 ACCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 314

```

```

Db 618 ACAATGATGATGATTATGTAATGATGATGAGAGATGAGAAATGAGATGATATG 559
Qy 315 TGGCGAACTGTACATGCTCACACCCAGAAAAACATGCTGACCTGGTGACCTCGTGACCT 374
Db 558 TGACAAATGGTAATGATGGTGAATGAGAGAGACATGCGGATGATGATGAAGATGATGA 499
Qy 375 GGTGACGATGATGAGGCGGTGTTAATGAAGTTCATCATTTATGCTGGTTGACATGA 434
Db 498 TGTGGCGATGAGAGATGAGATGAGAGAGAGGTTGATGATGATGATGATGATGATGA 439
Qy 435 TGGT 438
Db 438 TGGT 435

```

Search completed: November 29, 2002, 07:43:03
 Job time : 70 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2002, 10:09:03 ; Search time 2876 Seconds

(without alignments)
15087.720 Million cell updates/sec

Title: US-09-868-760-6

Sequence: 1 g9y9gggagcagtgtaacga.....gctgtatagagacggca 1491

Scoring table:

OLIGO_MUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:**
1: gb_ha:**
2: gb_hg:**
3: gb_in:**
4: gb_om:**
5: gb_ov:**
6: gb_pa:**
7: gb_ph:**
8: gb_pl:**
9: gb_pr:**
10: gb_ro:**
11: gb_sts:**
12: gb_sy:**
13: gb_un:**
14: gb_vl:**
15: em_ba:**
16: em_fun:**
17: em_hum:**
18: em_in:**
19: em_mu:**
20: em_om:**
21: em_or:**
22: em_ov:**
23: em_pat:**
24: em_ph:**
25: em_pl:**
26: em_ro:**
27: em_sts:**
28: em_un:**
29: em_vl:**
30: em_hg_hum:**
31: em_hg_inv:**
32: em_hg_inv:**
33: em_hg_mus:**
34: em_hg_pln:**
35: em_hg_rtd:**
36: em_hg_mam:**
37: em_hg_vrt:**
38: em_sy:**
39: em_hgo_hum:**
40: em_hgo_mus:**
41: em_hgo_other:**

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1422	95.4	1700	3	AF273766	AF273766 Perna can
2	26	1.7	10210	3	AC020131	AC020131 Drosophila
3	26	1.7	17735	3	AC005714	AC005714 Drosophila
4	26	1.7	307363	3	AE003457	AE003457 Drosophila
5	25	1.7	65259	2	AC108920	AC108920 Mus muscu
6	25	1.7	67759	2	AC109858	AC109858 Drosophila
7	25	1.7	192588	3	AC010057	AC010057 Drosophila
8	25	1.7	207018	2	AE003469	AE003469 Drosophila
9	25	1.7	302527	3	AE003469	AE003469 Drosophila
10	24	1.6	476	8	AF231797	AF231797 Pinnus tae
11	24	1.6	508	8	AF172408	AF172408 Pinnus tae
12	24	1.6	179289	10	AL590633	AL590633 Mouse DNA
13	24	1.6	209173	2	AL845279	AL845279 Mus muscu
14	22	1.5	1817	1	BRUBCSP	BRUBCSP
15	22	1.5	9960	2	AE009521	AE009521 Brucella
16	22	1.5	56956	2	AC117081	AC117081 Dictyoste
17	22	1.5	135691	2	AC095233	AC095233 Rattus no
18	22	1.5	156312	2	AC121174	AC121174 Rattus no
19	22	1.5	165864	2	AC123157	AC123157 Rattus no
20	22	1.5	168909	2	AC017470	AC017470 Drosophila
21	22	1.5	169289	3	AC007145	AC007145 Drosophila
22	22	1.5	265605	3	AE003562	AE003562 Drosophila
23	22	1.4	24	6	A20717	A20717 Oligonucleo
24	23	1.4	33	6	AL9463	AL9463 Oligonucleo
25	23	1.4	23	3	AF286896	AF286896 Plasmodiu
26	23	1.4	706	3	AF223572	AF223572 Hirudo ni
27	21	1.4	846	9	AB044395S4	AB044395 Homo sapi
28	21	1.4	1587	8	AY084494	AY084494 Arabidops
29	21	1.4	1965	3	DM073490	DM073490 Arabidops
30	21	1.4	2108	3	AF272790	AF272790 Hirudo ni
31	21	1.4	2922	3	AF071417	AF071417 Drosophila
32	21	1.4	2980	3	AY058737	AY058737 Drosophila
33	21	1.4	4194	8	MACR1	MACR1
34	21	1.4	13653	2	AC018218	AC018218 Drosophila
35	21	1.4	41789	3	CBRG44A05	CBRG44A05
36	21	1.4	82033	8	AB011476	AB011476 Arabidops
37	21	1.4	85862	3	AC004433	AC004433 Drosophila
38	21	1.4	104711	9	AL391988	AL391988 Human DNA
39	21	1.4	139581	2	AP005055	AP005055 Oryza sat
40	21	1.4	146388	2	AP005880	AP005880 ciltb.43.a
41	21	1.4	155152	2	AP004858	AP004858 Oryza sat
42	21	1.4	163012	3	AC008289	AC008289 Drosophila
43	21	1.4	168635	2	AC102289	AC102289 Mus muscu
44	21	1.4	171972	3	AC012388	AC012388 Drosophila
45	21	1.4	182278	2	AC121730	AC121730 Rattus no

ALIGNMENTS

RESULT 1
AF273766
LOCUS AF273766 1700 bp mRNA linear INV 20-MAR-2001
DEFINITION Perna canaliculus perlin precursor, mRNA, complete cds.
ACCESSION AF273766
VERSION AF273766.1 GI:13383377
KEYWORDS
SOURCE Perna canaliculus.
ORGANISM Perna canaliculus
Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
Mytiloidea; Mytilidae; Perna.
REFERENCE
1 (bases 1 to 1700)
Scotti,P.D., Dearling,S.C., Greenwood,D.R. and Newcomb,R.D.
Pernin: a novel, self-aggregating haemolymph protein from the New
Zealand green-lipped mussel, Perna canaliculus (Bivalvia).

Mytilidae)
JOURNAL Comp. Biochem. Physiol. B, Biochem. Mol. Biol. 128 (4), 767-779
(2001)
MEDLINE 21186417
PUBMED 11290439
2 (bases 1 to 1700)
REFERENCE Scott, P.D., Deating, S.C., Greenwood, D.R. and Newcomb, R.D.
AUTHORS Direct Submission
TITLE Submitted (31-May-2000) The Horticulture and Food Research
JOURNAL Institute of New Zealand Ltd, 120 Mt. Albert Road, Auckland, New
Zealand

FEATURES
source Location/Qualifiers
1..1700
/organism="Perna canaliculus"
/db_xref="taxon:38949"
34..1587
/note="haemolymph protein; N-terminus determined by
microsequencing of perlin: DGEQCDGON and HPLC-purified
cNbr and tryptic digest fragments: ASSLHHVHG; VVNEVHH;
GSHPEIYH; YHGHDRA; QGHEVESERVACVIGRA"
/evidence="experimental"
/product="perlin precursor"
/protein_id="AAK20952.1"
/db_xref="GI:1383378"
/translation="MKILLSLVYFALALQVADDEOCNDQNDKDDHDDHDDHND
DHDDETHNYACQEMEPNPMASLHHVHGSIELSQQHGVLELHLVGSFNSD
HDHNGHLLHMLGDSKSGDSISGLNAPENKADPGDLDVDDRELVNEVHTA
WDDIDETALFALIGHSMTILQSGHTDAPTRASRIACCVIGRAKRETNALHNL
EEDKETHANCDVNSNTHQPAKLNHHVHGTLDRQYGGVLESLHSEFNYSIDKID
HLHDVITANGDLTSGCDNLKADYDHEHSELDGIDHDDHGVNESHSYWIN
ITEDSVLGRSTAIHRODHLRSKIAICVIGQSHPEIYHAKVVPNESTGLH
HNVSGSTFEQPTGGSTHMTDLKGFNVSDLSHHRHGVLEHMGMSHGSISLGRY
HGHDADHDKRPGLDVIDSDSHGIVHSTRFPHLVNEDLMSLVIMQGHEVESER
VACCVIGRA"

CDS
sig_peptide
mat_peptide
polya_signal
BASE COUNT 515 a 375 c 379 g 431 t
ORIGIN

Query Match 95.4% Score 1422; DB 3; Length 1700;
Best Local Similarity 99.9% Pred. No. 0;
Matches 1472; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 19 GATGGGAGAAAGATGACACCATGACACACACCATGATGATGACGACCAT 78
DB 112 GATGGGAGAAAGATGACACCATGACACACACCATGATGATGACGACCAT 171
OY 79 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 138
DB 172 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 231
OY 139 GCTAGCAGCCTTACACACCATGATGATGATGATGATGATGATGATGATG 198
DB 232 GCTAGCAGCCTTACACACCATGATGATGATGATGATGATGATGATGATG 291
OY 199 GCTTTATCTAGAACTTCTTGTGCGATGATGATGATGATGATGATGATGAT 258
DB 292 GCTTTATCTAGAACTTCTTGTGCGATGATGATGATGATGATGATGATGAT 351
OY 259 CATCATGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 318
DB 352 CATCATGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 411
OY 319 GAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 378
DB 412 GAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 471
OY 379 GAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 438
DB 472 GAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 531

OY 439 ACAGACCAAAACACCGAAAGCTCTCATGACACTCAATGACTATTTTACAGAGAGTGC 498
DB 532 ACAGACCAAAACACCGAAAGCTCTCATGACACTCAATGACTATTTTACAGAGAGTGC 591
OY 499 ACCGATGCTGATACCCGACGAGTAGAATGCGCTGTGTGTTATGTCATGAGAAAGCT 558
DB 592 ACCGATGCTGATACCCGACGAGTAGAATGCGCTGTGTGTTATGTCATGAGAAAGCT 651
OY 559 CGCCGAGAAACAGACAGCTGCTCTCATGACAGTACAGAGTAGAGAGATGAAATGAGCATTTAT 618
DB 652 CGCCGAGAAACAGACAGCTGCTCTCATGACAGTACAGAGTAGAGAGATGAAATGAGCATTTAT 711
OY 619 GCCCATGCTGAGTAGAATGATGATGATGATGATGATGATGATGATGATGATGATG 678
DB 712 GCCCATGCTGAGTAGAATGATGATGATGATGATGATGATGATGATGATGATGATG 771
OY 679 GGAACCATGATTTTCAAAAGATTGATTGATGATGATGATGATGATGATGATGATG 738
DB 772 GGAACCATGATTTTCAAAAGATTGATTGATGATGATGATGATGATGATGATGATG 831
OY 739 GGAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 798
DB 832 GGAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 891
OY 799 GGTGACCTGACGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 858
DB 892 GGTGACCTGACGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 951
OY 859 CACAGTGAATGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 918
DB 952 CACAGTGAATGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1011
OY 919 AGCCAGATGATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 978
DB 1012 AGCCAGATGATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1071
OY 979 GCCATTCACCAAGAGACATCTTCATGATGATGATGATGATGATGATGATGATGATG 1038
DB 1072 GCCATTCACCAAGAGACATCTTCATGATGATGATGATGATGATGATGATGATGATG 1131
OY 1039 CGTGGACAGAGCATGCAAGAAATGTTTACAGAGAGTAAATGTTGTCAGACCTTAATGCA 1098
DB 1132 CGTGGACAGAGCATGCAAGAAATGTTTACAGAGAGTAAATGTTGTCAGACCTTAATGCA 1191
OY 1099 GAATCTACTGTTTACATCAGCATGCTCTGTTGTTTAAATGATGAAACAGACCTTGGGA 1158
DB 1192 GAATCTACTGTTTACATCAGCATGCTCTGTTGTTTAAATGATGAAACAGACCTTGGGA 1251
OY 1159 GGATCAACATATGAGCGGTGATGATGATGATGATGATGATGATGATGATGATGATG 1218
DB 1252 GGATCAACATATGAGCGGTGATGATGATGATGATGATGATGATGATGATGATGATG 1311
OY 1219 CATGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1278
DB 1312 CATGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1371
OY 1279 GGCAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1338
DB 1372 GGCAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1431
OY 1339 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1398
DB 1432 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1491
OY 1399 GTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1458
DB 1492 GTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1551
OY 1459 GAGAGGTTGCTTGTGTTATGAGAGGGCA 1491
DB 1552 GAGAGGTTGCTTGTGTTATGAGAGGGCA 1584

```

RESULT 2
LOCUS AC020131
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC020131
VERSION AC020131.1
KEYWORDS HTG: HTG_6664766
SOURCE Drosophila melanogaster.
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS Adams, M. and Venter, J.C.
JOURNAL Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10212152 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1. 10210
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 3082 a 2176 c 2398 g 2554 t
ORIGIN
Query Match 1.7%; Score 26; DB 2; Length 10210;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 922 CACGATATTCCTCGATCAATATCTT 947
|||||
Db 3128 CACGATATTCCTCGATCAATATCTT 3153
|||||

RESULT 3
LOCUS AC005714
DEFINITION Drosophila melanogaster, chromosome 2R, region 58D4-58E2, BAC clone
BACR48M13, complete sequence.
ACCESSION AC005714
VERSION AC005714.10
KEYWORDS HTG: HTG_16874830
SOURCE Drosophila melanogaster.
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS Celinker, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanalides, P.G., Brandon, R.C.,
Rogers, Y., An, H., Baldwin, D., Banazon, J., Beeson, K.Y., Busam, D.A.,
Caloson, J.W., Center, A., Champs, M., Davenport, L.B., Dietz, S.M.,
Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D.,
Fertler, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A.,
Gonzalez, M., Houck, J., Hoskins, R.A., Hostlin, D., Howland, T.J.,
Ibegwam, C., Jaitai, M., Kruse, D., Li, P., Mattei, B., Mostrefi, A.,
McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,
Phalen, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,
Poulsen, V., Pongas, S., Piltman, G.S., Puri, V., Richards, S., Schelker, F.,
Stebler, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M.,
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
JOURNAL Sequencing of Drosophila chromosome 2R, region 58D4-58E2
Unpublished
2 (bases 1 to 177735)
REFERENCE
AUTHORS Celinker, S.E., George, R.A., Galle, R., Svirskas, R.R., Hoskins, R.A.,
Aghayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Chavez, C.,

```

```

TITLE
JOURNAL Direct Submission
COMMENT Submitted (26-SEP-1998) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Nov 9, 2001 this sequence version replaced 91:4731051.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu.
FEATURES
source
1. 177735
/organism="Drosophila melanogaster"
/strain="Y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="2R"
/map="58D4-58E2"
/clone="BACR48M13 (D436)"
/clone_11b="RPCT-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBAC3.0)"
BASE COUNT 49614 a 39531 c 40152 g 48418 t
ORIGIN
Query Match 1.7%; Score 26; DB 3; Length 177735;
Best Local Similarity 100.0%; Pred. No. 0.0037;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 922 CACGATATTCCTCGATCAATATCTT 947
|||||
Db 99256 CACGATATTCCTCGATCAATATCTT 99281
|||||

RESULT 4
LOCUS AE003457
DEFINITION Drosophila melanogaster 2R section 65 of 74 of the complete arm,
complete sequence.
ACCESSION AE003457
VERSION AE003457.2
KEYWORDS HTG: HTG_21626551
SOURCE Fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS Adams, M.D., Celinker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,
Amanalides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.,
George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,
Sutton, G.G., Wortman, J.R., Vandeil, M.D., Zhong, Q., Chen, L.X.,
Brandon, R.C., Rogers, Y.H., Blazej, R.G., Champs, M., Pfeiffer, B.D.,
Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gaber
Miklos, G.L., Abriil, J.F., Aghayani, A., An, H.J.,
Andrews, P., Binkov, C., Baldwin, D., Ballew, R.M., Beeson, K.Y.,
Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y.,
Benos, P.V., Berman, B.P., Bhandari, P., Bolshakov, S., Borkova, D.,
Botchan, M.R., Bouck, J., Brokstein, P., Brothier, P., Butts, K.C.,
Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I.,
Cherry, J.M., Cawley, S., Dahle, C., Davenport, L.B., Davies, P., de
Pablos, K., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M.,
Dudman, P., Durbin, K.J., Evans, M., Dugan-Rocha, S., Dunlop, B.C.,
Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Fertler, S.,

```

[illegible]

db_xref="FLYBASE:FBan0003413"
/db_xref="FLYBASE:FBan0034718"
complement(join(11764..14560,25247..25467))
/gene="windpipe"
/product="CG3413-RD"
/db_xref="FLYBASE:FBan0003413"
/db_xref="FLYBASE:FBan0034718"
complement(join(11764..14560,21759..21807,24635..24857))
/gene="windpipe"
/product="CG3413-RC"
/db_xref="FLYBASE:FBan0003413"
/db_xref="FLYBASE:FBan0034718"
complement(join(11764..14560,21759..21807,23481..23501))
/gene="windpipe"
/product="CG3413-RA"
/db_xref="FLYBASE:FBan0003413"
/db_xref="FLYBASE:FBan0034718"
complement(12457..14490)
/gene="windpipe"
/note="windpipe gene product from transcript CG3413-RA"
/codon_start=1
/product="CG3413-PA"
/protein_id="AAM68219.1"
/db_xref="GI:21626553"
/db_xref="FLYBASE:FBan0003413"
/db_xref="FLYBASE:FBan0034718"
/db_xref="FLYBASE:FBan0034718"

Query Match 1.7%: Score 26; DB 3; Length 307363;
Best Local Similarity 100.0%; Pred. No. 0.0039;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 922 CACAGATATCTCGATCATATCTT 947
|||||
Db 134299 CACAGATATCTCGATCATATCTT 134324

RESULT 5
AC108920
LOCUS AC108920 65256 bp DNA linear HTG 14-JUN-2002
DEFINITION Mus musculus clone RP24-398E22, LOW-PASS SEQUENCE SAMPLING.
AC108920
AC108920.2 GI:21426206
HTG, HTGS_PHASE0.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,
Brown,A., Camarata,J., Campiano,A., Chang,J., Chazaro,B.,
Choquel,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardina,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Huine,W., Iliev,I., Johnson,R., Jones,C.,
Kamela,A., Karatas,A., Kells,C., Labocque,K., Lamazares,R.,
Lander,E., Lehoczy,J., Levine,R., Liu,G., Maclean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Melidim,J., Meneus,L., Mihova,T.,
Mlepan,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Poljara,R., Roman,J.,
Rozell,M., Roy,A., Santos,R., Schauer,B., Stange-Thomann,N.,
Severly,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,

TITLE
JOURNAL
COMMENT

REFERENCE
AUTHORS
Zainoun,J., Zemdek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 65256)
Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhalter,B., Brown,A., Camarata,J., Campiano,A., Chang,J.,
Chazaro,B., Choquel,Y., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardina,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Huine,W., Iliev,I., Johnson,R., Jones,C.,
Kamela,A., Karatas,A., Kells,C., Labocque,K., Lamazares,R.,
Lander,E., Lehoczy,J., Levine,R., Liu,G., Maclean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Melidim,J., Meneus,L., Mihova,T.,
Mlepan,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Poljara,R., Roman,J.,
Rozell,M., Roy,A., Santos,R., Schauer,B., Stange-Thomann,N.,
Severly,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zemdek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (14-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WMR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project information
Center project name: L21124
Center clone name: 398_E-22

NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 745: contig of 745 bp in length
746 845: gap of 100 bp
846 1599: contig of 754 bp in length
1600 1699: gap of 100 bp
1700 2445: contig of 746 bp in length
2446 2545: gap of 100 bp
2546 3287: contig of 742 bp in length
3288 3387: gap of 100 bp
3388 4099: contig of 712 bp in length
4100 4199: gap of 100 bp
4200 4950: contig of 751 bp in length
4951 5050: gap of 100 bp
5051 5787: contig of 737 bp in length
5788 5887: gap of 100 bp
5888 6636: contig of 749 bp in length
6637 6736: gap of 100 bp
6737 7473: contig of 737 bp in length
7474 7573: gap of 100 bp
7574 8322: contig of 749 bp in length

8323 8422: gap of 100 bp
8423 9168: contig of 746 bp in length
9169 9268: gap of 100 bp
9269 10024: contig of 756 bp in length
10025 10124: gap of 100 bp
10125 10856: contig of 732 bp in length
10857 10956: gap of 100 bp
10957 11693: contig of 737 bp in length
11694 11793: gap of 100 bp
11794 12530: contig of 737 bp in length
12531 12630: gap of 100 bp
12631 13358: contig of 728 bp in length
13359 13458: gap of 100 bp
13459 14189: contig of 731 bp in length
14190 14289: gap of 100 bp
14290 15015: contig of 726 bp in length
15016 15115: gap of 100 bp
15116 15856: contig of 741 bp in length
15857 15956: gap of 100 bp
15957 16699: contig of 743 bp in length
16700 16799: gap of 100 bp
16800 17548: contig of 749 bp in length
17549 17648: gap of 100 bp
17649 18382: contig of 734 bp in length
18383 18482: gap of 100 bp
18483 19204: contig of 722 bp in length
19205 19304: gap of 100 bp
19305 20056: contig of 752 bp in length
20057 20156: gap of 100 bp
20157 20896: contig of 740 bp in length
20897 20996: gap of 100 bp
20997 21742: contig of 746 bp in length
21743 21842: gap of 100 bp
21843 22571: contig of 729 bp in length
22572 22671: gap of 100 bp
22672 23410: contig of 739 bp in length
23411 23510: gap of 100 bp
23511 24251: contig of 741 bp in length
24252 24351: gap of 100 bp
24352 25091: contig of 740 bp in length
25092 25191: gap of 100 bp
25192 25940: contig of 749 bp in length
25941 26040: gap of 100 bp
26041 26788: contig of 748 bp in length
26789 26888: gap of 100 bp
26889 27619: contig of 751 bp in length
27620 27719: gap of 100 bp
27720 28356: contig of 617 bp in length
28357 28456: gap of 100 bp
28457 29185: contig of 749 bp in length
29186 29285: gap of 100 bp
29286 30025: contig of 740 bp in length
30026 30125: gap of 100 bp
30126 30880: contig of 755 bp in length
30881 30980: gap of 100 bp
30981 31713: contig of 733 bp in length
31714 31813: gap of 100 bp
31814 32567: contig of 754 bp in length
32568 32667: gap of 100 bp
32668 33400: contig of 733 bp in length
33401 33500: gap of 100 bp
33501 34251: contig of 751 bp in length
34252 34351: gap of 100 bp
34352 35103: contig of 752 bp in length
35104 35203: gap of 100 bp
35204 35945: contig of 742 bp in length
35946 36045: gap of 100 bp
36046 36777: contig of 732 bp in length
36778 36877: gap of 100 bp
36878 37626: contig of 749 bp in length
37627 37726: gap of 100 bp
37727 38471: contig of 745 bp in length
38472 38571: gap of 100 bp

38572 39315: contig of 744 bp in length
39316 39415: gap of 100 bp
39416 40152: contig of 737 bp in length
40153 40252: gap of 100 bp
40253 40988: contig of 736 bp in length
40989 41088: gap of 100 bp
41089 41840: contig of 752 bp in length
41841 41940: gap of 100 bp
41941 42702: contig of 762 bp in length
42703 42802: gap of 100 bp
42803 43561: contig of 759 bp in length
43562 43661: gap of 100 bp
43662 44398: contig of 738 bp in length
44400 44499: gap of 100 bp
44500 445220: contig of 721 bp in length
45221 45320: gap of 100 bp
45321 46062: contig of 742 bp in length
46063 46162: gap of 100 bp
46163 46904: contig of 742 bp in length

Query Match 1.7%: Score 25; DB 2; Length 65256;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Yy 68 ATGACGACCATGATGATGATGATGA 92
Db 8829 ATGACGACCATGATGATGATGATGA 8853
|||||

RESULT 6
AC019858/c
LOCUS
DEFINITION
AC019858
Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, In ordered
pieces.
AC019858
AC019858.1 GI:665039
VERSION
HTG: HTG5.PHASE2.
KEYWORDS
Drosophila melanogaster.
SOURCE
Drosophila melanogaster.
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 67759)
ADAMS, M. and VENTER, J. C.
AUTHORS
Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
TITLE
Direct Submission
JOURNAL
This sequence was identified as CDM:1021184 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
LOCATION/Qualifiers
1..67759
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

BASE COUNT 18034 a 15223 c 15520 g 18982 t
ORIGIN

Query Match 1.7%: Score 25; DB 2; Length 67759;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 68 ATGACGACCATGATGATGATGATGA 92
Db 33276 ATGACGACCATGATGATGATGATGA 33252
|||||

RESULT 7
AC010057
LOCUS
DEFINITION
AC010057
Drosophila melanogaster 3L BAC RP98-26C18 (Roswell Park Cancer

Identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases.

Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annot.action.html>.

FEATURES

source Location/Qualifiers
1..192588
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3L"
/clone="RP98-26C18"
BASE COUNT 52876 a 43443 c 42611 g 53658 t
ORIGIN

Query Match 1.7%; Score 25; DB 3; Length 192588;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 68 ATGACGACATGATGATGATGATGA 92
|||||
Db 171304 ATGACGACATGATGATGATGATGA 171328

RESULT 8
AC116539
LOCUS
DEFINITION
Drosophila melanogaster clone RP98-18E1, WORKING DRAFT SEQUENCE, 3
unordered pieces:
AC116539 207018 bp DNA linear HTG 23-JUL-2002
AC116538
AC116538 2 61-21930209
HTG: HTGS_PHASE1 HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster

Phylophila melanogaster
Eukaryota: Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 207018)
Muzny,D.M., Adams,C., Adio-Oduola,B., All-osman,F.R., Allen,C.,
Aisbrook,S.L., Amaralunga,H.C., Are,J.R., Ayala,M., Banks,T.,
Barbato,J., Benton,J., Bimaga,K., Blankenburg,K., Bonaldi,D.,
Bouch,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Degen,A.L., Ding,Y., Dinn,H.H.,
Doutswarte,K.J., Draper,H., Dugan-Rocha,S., Dublin,K.J.,
Ehrhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabriel,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorelli,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Hayak,P., Hayes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hughes,M., Hollway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolyvet,S., Joudan,S.,
Karlsson,B., Kelly,S., Khan,O., King,L., Kovach,C.,
Kraus,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lien,C., Liu,J., Liu,W., Louise,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Lune,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Melker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
JOURNAL
Worley,K.C.
Submitted (29-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 207018)
Worley,K.C.
Direct Submission
Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 23, 2002 this sequence version replaced g1:19807671.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: DRML
Center clone name: RP98-18E1
----- Summary Statistics
Chemistry: Dye-terminator Big Dye; Infinity% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 219317 bases at least Q40
Consensus quality: 224732 bases at least Q30
Consensus quality: 233117 bases at least Q20
Estimated insert size: 207648; sum-of-coverage estimation
Quality coverage: 10x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 6420: contig of 6420 bp in length
* 6421 6530: gap of unknown length
* 91457 91456: contig of 84936 bp in length
* 91457 91556: gap of unknown length
* 91557 207018: contig of 115462 bp in length.
----- Location/Qualifiers
1..207018
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="RP98-18E1"
BASE COUNT 5739 a 45880 c 45706 g 57469 t 224 others
ORIGIN

Query Match 1.7%; Score 25; DB 2; Length 207018;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```
/db_xref="FLYBASE:FBan0012189"
/db_xref="FLYBASE:FBgn0035150"
join(26843..26959,27018..28375,28430..28712,28775..29198,
29261..29314,29380..30067,30132..30455)
/gene="CG12189"
/product="CT9555"
/db_xref="FLYBASE:FBan0012189"
/db_xref="FLYBASE:FBgn0035150"
join(26924..26959,27018..28375,28430..28712,28775..29198,
29261..29314,29380..30067,30132..30276)
/gene="CG12189"
/note="CG12189 gene product"
/codon_start=1
/protein_id="AAF47401.1"
/db_xref="GI:7291986"
/db_xref="FLYBASE:FBan0012189"
/db_xref="FLYBASE:FBgn0035150"
/translation="MTREDNGFSEMGGEFAKSKLEQFAAASDPFRKSDLPFGIS
IFVNGRTDPSADELRKIMVHGTFHHYERSHTYIIASVLPVKRNNLSKFTSAK
WVVDLEKKRIYDYPVLYTNQTSQPMIFGPKDNGNESKSDVEPPKDAEVEY
DSTKDETOMELGILKNQAVATSPKESASASEKINTSTSNSTSTAAADPNF
LSEFYKNSRLHIATLGAQAVATSPKESASASEKINTSTSNSTSTAAADPNF
HIDMDFEVSGLATFRLRGLPIAVTHSGKNADVPVHPQADRAKELTFAQFE
HHPHDQKAEKRSQFDMKSLSEIASCSEARERKIRNGMEFGALKICPELKTIPY
DEPGYEVAFETLDVAVTNTIEASCDEFEVDLAEHLVDMVMAFVSHLRREVY
SKTGPCSGAVGNKLLAMATKEKPNQGLDSSNDILAEVMAFVSHLRREVY
SHKQAGLNGCGVONTLEKMEVGLKLGONLQNGIDDRPLAVEOIRKTVSA
EMNGIRFNSVECEQFQCOLSEVTRVETRRKRSIKLIGVRAAPAVETSKY
GHVCDIINKSLIKYATDVNLTITVLDMDADIPDELRGKIHITRLEDNEY
RKNNIKREMGKSEMRKDKIPQAGVDSIGDDYKPKVPEKPKPREPNVLSM
LTAAVRSKSVTEEDRSORGTSKPIITRPLSLIPKDEVDLAQLPEDIRLELVANREH
CIAEYISDOIVADYIDDLPPDPVPHIILKILSHPEMDPELMDGNYDLNDPVSERY
PKNDVDLLIKOVSRMINKDOLDHVCDVMKWKCRILIMKRSSCCNHVAKIIESIO
NOMLTREGSLTFEYIKICS"
/complement(44992..52784)
/gene="CG17129"
/note="CG17129"
/nup="61C8-61C9"
/db_xref="FLYBASE:FBan0017129"
/db_xref="FLYBASE:FBgn0035151"
/complement(join(4992..5007,3,50495..51310,51776..52309,
52379..52784))
/gene="CG17129"
/product="CT38022"
/db_xref="FLYBASE:FBan0017129"
/db_xref="FLYBASE:FBgn0035151"
/complement(join(4992..5007,3,50495..51310,51776..52227))
/gene="CG17129"
/note="CG17129 gene product"
/codon_start=1
/protein_id="AAF47402.1"
/db_xref="GI:7291987"
/db_xref="FLYBASE:FBan0017129"
/db_xref="FLYBASE:FBgn0035151"
/translation="MAKWLRLDEVEGEEKSOODKMESESPAOEOKOKVLANPSV
VDCSEVEKSVFRDMETLDVESVLPNSILVDDGETTODKADYDINKLDE
ASLRKRNREEDSPSHNAKORSIKELFTPKKSGODGTSFAROKLTKDVAENM
IRYRDLQEOVDSEPKALGTAKELRLIPKOCNLSIOPTTMDXONPVSVTST
NSPNEPDAPGITYTOINRLPSPAPATSSNPATPNSPVSOTSGILPVPF
HDENPPLSVTTSVTVINSRVPDPBPANATWYTPNHNPATVPVPTSTSTXOPT
SSSYGCONPVAVLSSGYIRGPPDPANNAAGISNPRDGLNTIRMLADVPFAYC
ONNREAYHIVNLRHSLOTGASNRFQAPSEMPAIKDKREKREKSLQDDANAFNPF"
```

Query Match 1.7%: Score 25; DB 3; Length 302527;

Best Local 25; Similarity 100.0%; Pred. No. 0.016;

Matches 0; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 ATGACGACCATGATGATGATGATGA 92

DB 197335 ATGACGACCATGATGATGATGATGA 197359

RESULT 10

```
AF231797
LOCUS AF231797 476 bp DNA linear
DEFINITION Pinus taeda clone LC15112 microsatellite sequence.
ACCESSION AF231797
VERSION AF231797.1 GI:15004298
KEYWORDS
SOURCE Pinus taeda.
ORGANISM Pinus taeda.
REFERENCE 1 (bases 1 to 476)
AUTHORS Elsik,C.G. and Williams,C.G.
TITLE Microsatellite organization in a large plant genome
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 476)
TITLE Elsik,C.G. and Williams,C.G.
REFERENCE Direct Submission
AUTHORS Elsik,C.G. and Williams,C.G.
JOURNAL Submitted (05-FEB-2000) Department of Forest Science, Texas A&M University, Room 305 HFSB, College Station, TX 77843-2135, USA
FEATURES
source
1..476
/organism="Pinus taeda"
/db_xref="taxon:3352"
/clone="LC15112"
repeat_region
1..476
/note="microsatellite: trinucleotide repeat"
/rpt_type=tandem
BASE COUNT 187 a 50 c 93 g 144 t 2 others
ORIGIN
1..476
/note="microsatellite: trinucleotide repeat"
/rpt_type=tandem
Query Match 1.6%: Score 24; DB 8; Length 476;
Best Local Similarity 100.0%; Pred. No. 0.033;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 69 TGACGACCATGATGATGATGATGA 92
DB 287 TGACGACCATGATGATGATGATGA 310
RESULT 11
LOCUS AF172408 508 bp DNA linear
DEFINITION Pinus taeda clone LC15112 microsatellite sequence.
ACCESSION AF172408
VERSION AF172408.1 GI:9719232
KEYWORDS
SOURCE Pinus taeda.
ORGANISM Pinus taeda.
REFERENCE 1 (bases 1 to 508)
AUTHORS Elsik,C.G. and Williams,C.G.
TITLE Microsatellite recovery from the low-copy component of a large plant genome
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 508)
TITLE Elsik,C.G. and Williams,C.G.
REFERENCE Direct Submission
AUTHORS Elsik,C.G. and Williams,C.G.
JOURNAL Submitted (27-JUL-1999) Forest Science, Texas A&M University, HFSB Room 305, College Station, TX 77843-2135, USA
FEATURES
source
1..508
/organism="Pinus taeda"
/db_xref="taxon:3352"
/clone="LC15112"
repeat_region
277..474
/note="microsatellite"
/rpt_type=tandem
/rpt_unit=tag
BASE COUNT 195 a 61 c 100 g 149 t 3 others
ORIGIN
1..508
Query Match 1.6%: Score 24; DB 8; Length 508;
```

Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 TGACGACCATGATGATGATGA 92
Db 286 TGACGACCATGATGATGATGA 309
|||||

RESULT 12
AL590633/c 179289 bp DNA linear ROD 05-APR-2002
LOCUS Mouse DNA sequence from clone RP23-211023 on chromosome X, complete
DEFINITION sequence.
ACCESSION AL590633
VERSION AL590633.8 GI:20068454
KEYWORDS HTG.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
Wall, M.
Direct Submission
Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquies@sanger.ac.uk Clone requests: clonequests@sanger.ac.uk
On Apr 7, 2002 this sequence version replaced gi:15617304.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30);
an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Emu, EMBL; SW,
SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
from the RPI-23 Mouse PAC library
constructed by the group of Pletier de Jong.
for further details see http://www.chori.org/bacpac/home.htm
VECTOR: pPAC3.0.

FEATURES
source
1..179289
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP23-211023"
/clone_11b="RPI-23"
BASE COUNT 56940 a 36892 c 36607 g 48850 t
ORIGIN

Query Match 1.6%; Score 24; DB 10; Length 179289;
Best Local Similarity 100.0%; Pred. No. 0.062;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 CACCATGACCATGATGATGAT 87
Db 175276 CACCATGACCATGATGATGAT 175253
|||||

RESULT 13
AL845279 209173 bp DNA linear HTG 09-AUG-2002
LOCUS Mus musculus chromosome X clone RP23-2E13, *** SEQUENCING IN
DEFINITION PROGRESS ***, 12 unordered pieces.

ACCESSION AL845279
VERSION AL845279.1 GI:22204974
KEYWORDS HTG; PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 209173)
McLay, K.
Direct Submission
Submitted (02-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquies@sanger.ac.uk Clone requests: clonequests@sanger.ac.uk
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquies@sanger.ac.uk
----- Project Information
Center project name: bm2E13
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 207186 bases at least Q40
Consensus quality: 207812 bases at least Q30
Consensus quality: 207812 bases at least Q20
Insert size: 208073; sum-of-contrigs
Quality coverage: 9.32x in Q20 bases; sum-of-contrigs quality
coverage: 9.44x in Q20 bases; agarose-tp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 11646: contrig of 11646 bp in length
* 11647 11746: gap of 100 bp
* 11747 27085: contrig of 15339 bp in length
* 27086 27185: gap of 100 bp
* 27186 30891: contrig of 3706 bp in length
* 30892 30991: gap of 100 bp
* 30992 62508: contrig of 31517 bp in length
* 62509 62608: gap of 100 bp
* 62609 68419: contrig of 5811 bp in length
* 68420 68519: gap of 100 bp
* 68520 126394: contrig of 57875 bp in length
* 126395 126494: gap of 100 bp
* 126495 142463: contrig of 15969 bp in length
* 142464 142563: gap of 100 bp
* 142564 145884: contrig of 3321 bp in length
* 145885 145984: gap of 100 bp
* 145985 194833: contrig of 48849 bp in length
* 194834 194933: gap of 100 bp
* 194934 198509: contrig of 3576 bp in length
* 198510 198609: gap of 100 bp
* 198610 206634: contrig of 8025 bp in length
* 206635 206734: gap of 100 bp
* 206735 209173: contrig of 2439 bp in length.
Location/Qualifiers
1..209173
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP23-2E13"
/clone_11b="RPI-23"
1..11646
/note="assembly_fragment:02548.0"
11747..27085
/note="assembly_fragment:02548.1"

misc_feature 27186..30891
/note="assembly-fragment:02412
fragment_chain:1"
misc_feature 30992..62508
/note="assembly-fragment:02521
fragment_chain:1"
misc_feature 62609..68419
/note="assembly-fragment:02028
fragment_chain:1"
misc_feature 68520..126394
/note="assembly-fragment:00458
fragment_chain:1"
misc_feature 126495..142463
/note="assembly-fragment:00628
fragment_chain:1"
misc_feature 142564..143884
/note="assembly-fragment:02995
fragment_chain:1"
misc_feature 145985..194833
/note="assembly-fragment:03312
fragment_chain:1"
misc_feature 194934..198509
/note="assembly-fragment:01240
fragment_chain:1"
misc_feature 198610..206634
/note="assembly-fragment:00991
fragment_chain:1"
misc_feature 206735..209173
/note="assembly-fragment:01353
fragment_chain:1"
clone_end:SP6
vector_side:right"
BASE COUNT 62758 a 43215 c 42828 g 59259 t 1113 others
ORIGIN
Query Match 1.6%; Score 24; DB 2; Length 209173;
Best Local Similarity 100.0%; Pred. No. 0.063;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 64 CACCATGACGACATGATGATGAT 87
|||||
Db 192630 CACCATGACGACATGATGATGAT 192653
RESULT 14
BRUCSP/C 1817 bp DNA linear BCT 26-APR-1993
LOCUS B.abortus BSCP31 gene encoding a 31-KDa cell surface protein,
DEFINITION complete cds.
ACCESSION M20404
VERSION M20404.1 GI:144104
KEYWORDS cell surface protein.
SOURCE B.abortus (strain 19) DNA, clone PBA31-R7.
ORGANISM Brucella melitensis biovar Abortus
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Brucellaceae; Brucella.
REFERENCE 1 (bases 1 to 1817)
AUTHORS Mayfield,J.E., Bricker,B.J., Godfrey,H., Crosby,R.M., Knight,D.J.,
Halling,S.M., Balinsky,D. and Tabatabai,L.B.
TITLE The cloning, expression, and nucleotide sequence of a gene coding
for an immunogen Brucella abortus protein
JOURNAL Gene 63 (1), 1-9 (1988)
MEDLINE 88255848
PUBMED 3133283
FEATURES
source location/Qualifiers
1..1817 /organism="Brucella melitensis biovar Abortus"
/db_xref="taxon:235"
/note="Biovar Abortus"
565..1354
/note="cell surface protein precursor"
/codon_start=1
/transl_table=11

/protein_id="AA22993.1"
/db_xref="GI:144105"
/translation="MKFGSKIRRLAAVAAGNALGASFAVAQAPTFEIRIGTGCTAGT
YPIIGGLIANAISGAGEKGVPGIVATVAASNSGVANIKALSGALSGEFGQSDVAWA
YNGTGLDYGKVEDIRLRLATLYPEIRIHVARAKNIKSVYADLKGKRSIDEPGSGTI
VDARIVLEAVGLTEDDIIKAEBHLKPPGAGRLKDGALDLYFPVGGYPTGATSELAISNG
ILSVITSGPRADKILIEKYSFSPSKDVPAGAKVDAPEPTTALVAADQVTSKOPDDLIY
NITKVLNEDTRKALDAGHAKKGLIKLDSATSSGIPHPGAEFFYEACVLK"
sig_peptide 565..648
mat_peptide 649..1551
/product="cell surface protein"
BASE COUNT 382 a 468 c 539 g 428 t
ORIGIN 515 bp upstream of HindIII site.
Query Match 1.5%; Score 22; DB 1; Length 1817;
Best Local Similarity 100.0%; Pred. No. 0.63;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1331 ACCTGTGATGTATAGATGA 1352
|||||
Db 1406 ACCTGTGATGTATAGATGA 1385
RESULT 15
AE009521/c 9960 bp DNA linear BCT 28-DEC-2001
LOCUS Brucella melitensis strain 16M chromosome I, section 78 of 195 of
DEFINITION the complete sequence.
ACCESSION AE009521 AE008917
VERSION AE009521.1 GI:17982737
KEYWORDS
SOURCE Brucella melitensis.
ORGANISM Brucella melitensis
Bacteria; Proteobacteria; alpha subdivision; Rhizobiales group;
Brucellaceae; Brucella.
REFERENCE 1 (bases 1 to 9960)
AUTHORS Delvecchio,V.G., Redkar,R.J., Patra,G. and Mujer,C.
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (1), 443-448 (2002)
PUBMED 11756688
REFERENCE 2 (bases 1 to 9960)
AUTHORS Delvecchio,V.G., Redkar,R.J., Patra,G. and Mujer,C.
JOURNAL Submitted (13-NOV-2001) Institute of Molecular Biology and
Medicine, University of Scranton, Scranton, PA 18510, USA
3 (bases 1 to 9960)
REFERENCE 3 (bases 1 to 9960)
AUTHORS Elzer,P.H. and Hagius,S.
JOURNAL Submitted (13-NOV-2001) Department of Veterinary Science, LSU Ag
Center, 111 Dairymple Building, Baton Rouge, LA 70803, USA
4 (bases 1 to 9960)
REFERENCE 4 (bases 1 to 9960)
AUTHORS Kaprali,V., Los,T., Ivanova,N., Anderson,I., Bhattacharya,A.,
Lykdis,A., Reznik,G., Jablonski,L., Larsen,N., D'Souza,M.,
Bernal,A., Mezur,M., Goltsman,E., Selkov,E., Haselkorn,R.,
Kypriides,N. and Overbeek,R.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2001) Integrated Genomics, Inc., 2201 W. Campbell
Park Drive, IL 60612, USA
5 (bases 1 to 9960)
REFERENCE 5 (bases 1 to 9960)
AUTHORS Letesson,J.-U.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2001) Unite de Recherche en Biologie Molculaire,
Laboratoire d'Immunologie et de Microbiologie, Universite of Namur,
61 rue de Bruxelles, Namur 5000, Belgium
6 (bases 1 to 9960)
REFERENCE 6 (bases 1 to 9960)
AUTHORS O'Callaghan,D.

TITLE Direct Submission
JOURNAL Submitted (13-NOV-2001) Faculte de Medecine, INSERM U431, Avenue
Kennedy, Nimes 30900, France
FEATURES Location/Qualifiers
source
1. .9960
/organism="Brucella melitensis"
/strain="16M"
/db_xref="taxon:29459"
/chromosome="I"
104. .1093
/gene="BMEI0796"
104. .1093
/gene="BMEI0796"
104. .1093
/gene="BMEI0796"
/codon_start=1
/transl_table=11
/product="31 kDa IMMUNOGENIC PROTEIN PRECURSOR"
/protein_id="AAL51977.1"
/db_xref="GI:17982738"
/transl_table=11
/translation="MKFGSKRLRLAANAAGATAGASFAVAQAFTPEFRITGCTACT
YTPIGGLANALISGGERGVPGLVATIAVSSNGSVANINAKSGALESGFTOSDVAVA
VTGVLGDKRGVLEDDILRLATLYPETIHVARADNIKSVADLKGRVSLDEPSGTI
VARIIVLSEVGLTEDDIKAEHLKPGAGEKLGADLDAFFVGGYPTGATISLAISSG
ISLVISGPEADKILEKYSFESKDVVAGAYKQVATPTLAAVAQWPTSAQPDLLIY
NTKVLNMEDEKALDAGHAKGLIKLDSATSSLGIPLPAGEFEYKEAGVLK"
1330. .3378
/gene="BMEI0797"
1330. .3378
/gene="BMEI0797"
/codon_start=1
/transl_table=11
/product="TRANSPORTER"
/protein_id="AAL51978.1"
/db_xref="GI:17982739"
/translation="MTEBONAKLAPMEIDEVKARELEKEEKFDESHFRPLAVNARIY
TLILSLFHYTAGFGLPBMTHRIGLHAFVGLVETFPFSRKGVDEPKPKRLITG
IGISVDMGLATIAVAVIHVPLIPDLDAFRGNTSTDVVLSLLILILETFRS
VGMPLITISVLEMLYALXGSMGILVHPATNSOLVHLITGCTGATGAGVAVY
VHFVLRGVPATRGIGOLFDCAAWAGRPAGAPAKVYSISGALFGYISGVANV
TVGSLTTPAMTRICRGYRPAAPVPSASSTGCTTPIMAAAFIMTEFLNPTITIL
AIIVAPFMHFEVYLAQVHFKAKTGLRGMTREMPLEKALKMDPITILPVLAVLI
LSCGTPYLAATNGTILCAVGLMFRKMTIGEVFGLRQKATLALAGAAGATVGLI
VGVTLTGVRKISTIVSTRAQDLPFTGTSVSWFAFQTLTFLTLVMAKATVLI
CGGITFANILIMATIAAPALGLVBPVIAHFVYVGLADITPPALAAVAAAG
ACADPFKGTGAFRLGKAVLPVPEVSPSLVLSNMPDEFIATFGCVIGITLAL
GAISGFFLVTRKIMENVLLIFAMLLVABEYSSIVGLILLPVVRLHVASRRPAP
"3541. .3927
/gene="BMEI0798"
3541. .3927
/gene="BMEI0798"
/codon_start=1
/transl_table=11
/product="Hypothetical Protein"
/protein_id="AAL51979.1"
/db_xref="GI:17982740"
/translation="MLSVEOVAMEIVQIRISSVGVKLYMEFVTEGSERTIVRIEND
TKELRDEIVIRRAAKITGDMAGACEGIEPDSLILTRPSARRAGDALELRQLDGS
LEDTPASDPVSVTSASIPASADPKS"
complement(4034. .6226)
/gene="BMEI0799"
complement(4034. .6226)
/gene="BMEI0799"
/RC_number="5.4.99.2"
/codon_start=1
/transl_table=11
/product="METHYLMALONYL-COA MUTASE"
/protein_id="AAL51980.1"
/db_xref="GI:17982741"
/translation="WRCLADPOGASIMDDGDEPTKTRADWETLAEKELRPADSLVW
HPRGIDVKKPLITODDLGIGLGLTLPGEFPLRGRAIVAGRPWTIRQAGSTAE
ESNAETRALAAGQGVVAADLATHRGIDHPRGEGVAGVADISVEDMKILDP
GIPLEKISVMTNGNAVILPILANIVAGEQGVPRVGLSGTINDILKEFVNRVTYIY
PPEPMRIADIIVATAOEMPKRNSISISGVMHOGAGATLVQELAFTLADREVRAA

LKGLVNDPFCRLSFFPAIGNFMEITAKIARAALLMSINKEDPKKPGSLMURTH
CQTSVNSLOEDOPYNNIVATFAEASALAGTOSLHTNSDEDAIALPTFSFARIARNT
QILIONETGQVTRVDPPLAGSYIESTINTELAKEMALIEFVETLQAGMTAABSGLPKR
LIEBAATRRQVADKGEVEIVGVNKFLEDEDEIDILEIDNSVARSQJARIANTRKEM
RNKOYEAALAELEKARSGEENLLAAVEASRARATVGEISDPMARAGDIAAAPKY
VKRYGTAAKDEPEYQTLVDRLGEFAKQAGKPPVWAKLGQDHRGAKIILASFGD
IGFEVLAPLPQTPDEADMAIKARVHVLGVSSILAAGKRTLLPOLVETLRRKGABNII
VVCQGVPRDYOYLEFDHGVAAVFGPGTINVLDICARAVLDMECKLRNL"
complement(6233. .8236)
/gene="BMEI0800"
complement(6233. .8236)
/gene="BMEI0800"
/RC_number="6.4.1.3"
/codon_start=11
/transl_table=11
/product="PROPYLONYL-COA CARBOXYLASE ALPHA CHAIN,
MITOCHONDRIAL"
/protein_id="AAL51981.1"
/db_xref="GI:17982742"
/translation="MIRKTLIANRGEIACRVIRKAKKGIATVAVYSADARNALHVKM
ADEVHIGPAPSNOSTIVIDKILAIKEIGADAVHPGGLSENPRFAALKAANVT
IGEPVNAITAMGDKITSKRLAEAGVSTPGMGLIEDADEAVRIAGSITGPMIKAS
AGGGGGMRLAMNDEAREGFQLSNEKSSFGDRIETKFTQPRHILEIQLDQH
GNVYVYGERCSIORRNQVIEAPSPLEDEATFKKAGBOAVLAKARAVYSGATVEF
IVQNNRFFYLEMNTRLQVEHPVETLITGIDLVEMIRVAGSKLAKARAVYSGATVEF
IESRLAEDPYRNFLPSIGRLTRYRPVEGHNMPGTIVRNDTGVFEFGGKISAYDPMI
AKLTWGPDRISALIDAMGHALDAFEVEGIGNLFLFVANDHPRFREGALTAFLAE
YPDGSGVKSDEDDARTLAAVAEINLVAORROTOIGRSLSPKHSIANDVNTLDGY
SLPVARIAEGSGTTINFIDGSLPIASDMHQSGLSGFTGKPLIAVKYSRSGTMR
RMGMDEVAVHVRKPRVAELAKLMPVKLPDPTSKLLCPMGVITTSILVADGETVAGO
PLATVEAMKENVLRAERATVKKRTTAGSGLAVDELLIMEF"
complement(8321. .9853)
/gene="BMEI0801"
complement(8321. .9853)
/gene="BMEI0801"
/RC_number="6.4.1.3"
/codon_start=11
/transl_table=11
/product="PROPYLONYL-COA CARBOXYLASE BETA CHAIN"
/protein_id="AAL51982.1"
/db_xref="GI:17982743"
/translation="MOELBQLEARRAARLGGCORIDAGHKGKLTARERIEVLLD
EGSFEEDDMVTHRCTDFEMAGQKVAGGVVTGNGTNGKQVVFSDPTVLGSLSE
THAKICKIMDAVKNCAVIGLINDSGYRIVTQIEGVADVAFRRNDASGVAPQI
SVINGPAGAGVSPAMTDFIPWVDSYMRVTGPDVVKVTVNEIVTAELELGASTHT
KKSVDAGVENDIEALEQVRLTFDFLELNNREKPPVPIFDPPARLEMLDTLAVDS
ATKPYDMKELILADDEGDFEIQEAFKNIITGFIIMDGOYGVVANOPMVLAQCLD
IDASRKAAVRFRCDAFNIPILTLVDVGFGLPTAGQYGVVIRHGAKLFLAVSQATP
MVTILTRKAVGAVDVAMSKHIGADINAMPTAIIAVMGAKGTEILYRSELDGEKI
AARTKEVEERFANPEVAERGFIDEVIMPHSSRRIRARAPASLRNQSGAPMKKHDTI
PL"

Search completed: November 29, 2002, 12:42:25
Job time : 5648 secs

BASE COUNT 1917 a 2822 c 2934 g 2287 t
ORIGIN
Query Match 1.5%: Score 22; DB 1; Length 9960;
Best local similarity 100.0%: Pred. No. 0.77;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1331 ACCTGGGATGATTAAGATCA 1352
|||||
DB 945 ACCTGGGATGATTAAGATCA 924
|||||

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2002, 08:14:04 ; Search time 270 Seconds
(without alignments)
12436.035 Million cell updates/sec

```
Title: US-09-868-760-6
Perfect score: 1491
Sequence: 1 gayyggagcagttaacga.....gctgltatagcagcgca 1491
```

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

word size :

Total number of hits satisfying chosen parameters: 4370478

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Listing first 45 summaries

Database :

N_Geneseq.010002.*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001C.DAT.*

SUMMARIES

Result	No.	Score	Match	Query length	ID	Description
1	1491	100.0	1491	21	AAA47150	DNA encoding a ser
2	1491	100.0	1611	21	AAA47151	DNA encoding a ser
3	26	1.7	3211	23	ABU06474	Drosophila melanog
4	26	1.7	6398	23	ABU10542	Drosophila melanog
5	21	1.7	2171	23	ABU23878	Drosophila melanog
6	21	1.4	1599	23	AAC39999	Arabidopsis thalia
7	21	1.4	3309	23	ABL5963	Drosophila melanog
8	21	1.4	4909	26	AAV18924	S
9	21	1.4	6699	23	ABL15962	Drosophila melanog

C	1	c	11	10	21	1.4	40023	24	ABL51954		Human solute carrier
			12	20	20	1.3	65	22	AAH45894		C parvum coding se
			13	20	20	1.3	461	24	ABL93659		Arabidopsis thalia
			14	20	20	1.3	706	21	AAC49072		Arabidopsis thalia
			15	20	20	1.3	708	21	AAC34259		Arabidopsis thalia
			16	20	20	1.3	879	23	AA67497		DNA encoding novel
			17	20	20	1.3	1774	21	AAC42052		Arabidopsis thalia
			18	20	20	1.3	2177	23	ABL25286		Arabidopsis thalia
			19	19	19	1.3	54	21	AAA60434		Drosophila melanog
			20	19	19	1.3	42	22	NM45895		Plasmod falcipar con
			21	19	19	1.3	87	19	AAV47595		C parvum coding se
			22	19	19	1.3	104	17	AAAT33359		Rhodococcus thaloe
			23	19	19	1.3	104	22	ABA72619		Ich-2 gene 5' prim
			24	19	19	1.3	104	22	ABA73973		Human foetal liver
			25	19	19	1.3	104	22	ABA38329		Human foetal liver
			26	19	19	1.3	104	22	ABA39059		Probe #16795 for g
			27	19	19	1.3	104	22	AAK21048		Probe #17525 for g
			28	19	19	1.3	104	22	AAK22426		Human brain expres
			29	19	19	1.3	104	22	AAK47200		Human brain expres
			30	19	19	1.3	104	22	AAK48593		Human bone marrow
			31	19	19	1.3	104	22	AAI26272		Human bone marrow
			32	19	19	1.3	104	22	AAI53036		Probe #16205 for g
			33	19	19	1.3	104	22	AAI54422		Probe #21722 used
			34	19	19	1.3	104	24	ABS21403		Probe #2108 used
			35	19	19	1.3	104	24	ABS22354		Human genome-deriv
			36	19	19	1.3	142	22	ABA73028		Human foetal liver
			37	19	19	1.3	142	22	AAK21460		Human brain expres
			38	19	19	1.3	142	22	AAK47621		Human bone marrow
			39	19	19	1.3	179	20	AAI53454		Probe #22140 used
			40	19	19	1.3	213	21	AAAI9099		Oropouche NP cDNA
			41	19	19	1.3	240	21	AAZ69097		multicloning site
			42	19	19	1.3	240	21	AAZ61540		DNA fragment of PG
			43	19	19	1.3	432	21	AAZ61541		Nucleotide sequenc
			44	19	19	1.3	432	21	AAZ34414		DNA sequence of N6
			45	19	19	1.3	465	22	ABL94190		Arabidopsis thalia
									ABL10206		Human breast cancer
ALIGNMENTS											
RESULT 1											
AAAA7150											
ID	AAAA7150	standard; DNA; 1491 BP.									
XX	AAAA7150;										
XX	03-OCT-2000	(first entry)									
XX	DNA encoding a serine protease inhibitor protein.										
DE	Serine protease inhibitor; green-flipped mussel; anti-thrombin;										
KW	divalent metal cation binding activity; dietary supplement;										
KW	anticoagulant; ss.										
XX											
OS	Perna canaliculus.										
XX											
PN	WO2000039165-A1.										
XX											
PD	06-JUL-2000.										
XX											
PF	23-DEC-1999;	99MO-NZ00227.									
XX											
PR	23-DEC-1998;	98NZ-0333568.									
PR	23-JUL-1999;	99NZ-0336906.									
XX											
FA	(HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.										
XX											
FI	Scotti PD, Dearing SC, Greenwood DR, Newcomb RD;										
XX											
XX	WPI: 2000-452375/39.										
DR	P-PSDB: AAY93750.										
XX											

PT New Perna canaliculus serine protease inhibitor protein exhibiting
PT anti-thrombin activity and divalent metal cation binding activity,
PT useful as an anticoagulant agent and as a dietary supplement -
XX
PS Claim 10; Page 9-10; 44pp; English.
XX
CC The present sequence encodes a serine protease inhibitor
CC protein. The protein is isolated from the green-lipped mussel
CC (Perna canaliculus), and exhibits, inter alia, anti-thrombin activity
CC and divalent metal cation binding activity. The serine protease
CC inhibitor protein has a molecular weight of about 55 kilo Daltons.
CC The protein, and its fragments, are useful in medicaments, in food,
CC as dietary supplements or as bioremediation agents. In the dietary
CC supplements, the protein is associated with or bound to at least one
CC divalent cation (such as calcium, magnesium or zinc) of dietary
CC significance. The proteins or their fragments are also useful as
CC anticoagulant agents.
XX
SQ Sequence 1491 BP; 428 A; 333 C; 350 G; 379 T; 1 other;
Query Match 100.0%; Score 1491; DB 21; Length 1491;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAYGGGAGCAGTGTACAGTGGGAGAAACAAAGATGACGACGACGACGAT 60
DB 1 GATGGGAGCAGTGTACAGTGGGAGAAACAAAGATGACGACGACGACGAT 60
QY 61 GATCACCATGACGACCATGATGATGATGATGATGATGATGATGATGATG 120
DB 61 GATCACCATGACGACCATGATGATGATGATGATGATGATGATGATGATG 120
QY 121 GAACCAACCCCTCATATGAGCTATGACGCTTACGACCATGTCATGACAT 180
DB 121 GAACCAACCCCTCATATGAGCTATGACGCTTACGACCATGTCATGACAT 180
QY 181 TCACAGAAAGGCTCATGAGAGCTGTTATCTAGAACCTCATCTTGGATTC 240
DB 181 TCACAGAAAGGCTCATGAGAGCTGTTATCTAGAACCTCATCTTGGATTC 240
QY 241 GAAGACCATGACGACCATGATGATGATGATGATGATGATGATGATGATG 300
DB 241 GAAGACCATGACGACCATGATGATGATGATGATGATGATGATGATGATG 300
QY 301 GATTGTGATCTTATGAGCACTGATGACATGTCACACGACGACGACGACG 360
DB 301 GATTGTGATCTTATGAGCACTGATGACATGTCACACGACGACGACGACG 360
QY 361 GACCTGGTACCTGTTAGCATGATGATGAGGCGTGGTTATGAGTTCTAT 420
DB 361 GACCTGGTACCTGTTAGCATGATGATGAGGCGTGGTTATGAGTTCTAT 420
QY 421 TGGTGGACATTTGATGATGATGATGATGATGATGATGATGATGATGATG 480
DB 421 TGGTGGACATTTGATGATGATGATGATGATGATGATGATGATGATGATG 480
QY 481 ATTTTACAAAGGAGTACACGCGATGATGATGATGATGATGATGATGATG 540
DB 481 ATTTTACAAAGGAGTACACGCGATGATGATGATGATGATGATGATGATG 540
QY 541 ATTGGTCATGGAAGAGCTGCGCAAGAAACAGACGCTCTACATCAGCATG 600
DB 541 ATTGGTCATGGAAGAGCTGCGCAAGAAACAGACGCTCTACATCAGCATG 600
QY 601 GATTAAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 660
DB 601 GATTAAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 660
QY 661 CTTCATCATCATGATGATGATGATGATGATGATGATGATGATGATGATG 720
DB 661 CTTCATCATCATGATGATGATGATGATGATGATGATGATGATGATGATG 720
QY 721 GTGTCTACCATTTAGAGGATTTATGATGATGATGATGATGATGATGATG 780

DB 721 GTGTCTACCATTTAGAGGATTTATGATGATGATGATGATGATGATGATG 780
QY 781 GTACAGATCTACGCAACGCTGACCTGACCATGATGATGATGATGATGATG 840
DB 781 GTACAGATCTACGCAACGCTGACCTGACCATGATGATGATGATGATGATG 840
QY 841 GATCCTCATGAAATTAACACAGTATGATGATGATGATGATGATGATGATG 900
DB 841 GATCCTCATGAAATTAACACAGTATGATGATGATGATGATGATGATGATG 900
QY 901 CATGCGCTTGTCAATGAAGAACCCACAGATATTCCTGATCAATATCTCG 960
DB 901 CATGCGCTTGTCAATGAAGAACCCACAGATATTCCTGATCAATATCTCG 960
QY 961 GTCTGGACCTTCTATGATGATGATGATGATGATGATGATGATGATGATG 1020
DB 961 GTCTGGACCTTCTATGATGATGATGATGATGATGATGATGATGATGATG 1020
QY 1021 GCTGTGTGTCATGAGACGCTGGAGACAGACCATCCAGAAATTTGTCAG 1080
DB 1021 GCTGTGTGTCATGAGACGCTGGAGACAGACCATCCAGAAATTTGTCAG 1080
QY 1081 GCTGTGACCTTAATACAGAAATCTACTGTTACATCAACCATGCTCTG 1140
DB 1081 GCTGTGACCTTAATACAGAAATCTACTGTTACATCAACCATGCTCTG 1140
QY 1141 TTGCAACGACCCCTGGAGATCAACATATGACGCTGATCTCAAAAGATT 1200
DB 1141 TTGCAACGACCCCTGGAGATCAACATATGACGCTGATCTCAAAAGATT 1200
QY 1201 ACTGAGACCTTGTCAATCAATCAATGATGATGATGATGATGATGATGAT 1260
DB 1201 ACTGAGACCTTGTCAATCAATCAATGATGATGATGATGATGATGATGAT 1260
QY 1261 CATGCGCTGACCTTCTAGGACAGAAATGATGATGATGATGATGATGAT 1320
DB 1261 CATGCGCTGACCTTCTAGGACAGAAATGATGATGATGATGATGATGAT 1320
QY 1321 AGACCTGGTACCTTGTGATGATGATGATGATGATGATGATGATGATG 1380
DB 1321 AGACCTGGTACCTTGTGATGATGATGATGATGATGATGATGATGATG 1380
QY 1381 ACCCTTGTATCTTATGATGATGATGATGATGATGATGATGATGATGAT 1440
DB 1381 ACCCTTGTATCTTATGATGATGATGATGATGATGATGATGATGATGAT 1440
QY 1441 GGACATGAGTGCAGAGTGAAGAGGTTGCTGCTGTTATGAGACGGGCA 1491
DB 1441 GGACATGAGTGCAGAGTGAAGAGGTTGCTGCTGTTATGAGACGGGCA 1491

RESULT 2
AAA47151
ID AAA47151 standard; DNA: 1611 BP.
XX
XX AAA47151;
DT 03-OCT-2000 (first entry)
XX
DE DNA encoding a serine protease inhibitor protein.
XX
KW Serine protease inhibitor; green-lipped mussel; anti-thrombin;
KW divalent metal cation binding activity; dietary supplement;
KW anticoagulant; ss.
XX
OS Perna canaliculus.
XX
FH Key Location/Qualifiers
FT CDS 1..1494
FT /tag= a
FT /product= "serine protease inhibitor"
FT polyA_site 1557..1563

```
FT      /*tag= b
XX      MO200039165-A1.
XX      PD      06-JUL-2000.
XX      PE      23-DEC-1999; 99MO-N200227.
XX      PR      23-DEC-1998; 98NZ-0333568.
XX      PR      23-JUL-1999; 99NZ-0336906.
XX      (HOKT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.
XX      PI      Scotti PD, Dearing SC, Greenwood DR, Newcomb RD;
XX      DR      WPI; 2000-452375/39.
XX      DR      P-PDB: AAT93750.
XX      PT      New Perna canaliculus serine protease inhibitor protein exhibiting
XX      PT      anti-thrombin activity and divalent metal cation binding activity,
XX      PT      useful as an anticoagulant agent and as a dietary supplement -
XX      Claim 11: Page 10-11: 44pp; English.
XX      CC      The present sequence encodes a serine protease inhibitor
XX      CC      protein. The protein is isolated from the green-lipped mussel
XX      CC      (Perna canaliculus), and exhibits, inter alia, anti-thrombin activity
XX      CC      and divalent metal cation binding activity. The serine protease
XX      CC      inhibitor protein has a molecular weight of about 55 kilo Daltons.
XX      CC      The protein, and its fragments, are useful in medicaments, in food,
XX      CC      as dietary supplements or as bioremediation agents. In the dietary
XX      CC      supplements, the protein is associated with or bound to at least one
XX      CC      divalent cation (such as calcium, magnesium or zinc) of dietary
XX      CC      significance. The proteins or their fragments are also useful as
XX      CC      anticoagulant agents.
XX      SQ      Sequence 1611 BP; 499 A; 348 C; 360 G; 402 T; 2 other:
XX
Query Match      100.0%; Score 1491; DB 21; Length 1611;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 GAYGGGAGCAGTGTAGAGTGGGAGAGAAAGATGACGACGATGACGACGACGAT 60
DB      1 GAYGGGAGCAGTGTAGAGTGGGAGAGAAAGATGACGACGATGACGACGACGAT 60
QY      61 GATCACCATGAGGACCATGATGATGATGATGATGATGATGATGATGATGATGAT 120
DB      61 GATCACCATGAGGACCATGATGATGATGATGATGATGATGATGATGATGATGAT 120
QY      121 GAACCAAAACCCCTCATATGAGCTAGACAGCCTTCACCCATGTCATGAGCATAGATTG 180
DB      121 GAACCAAAACCCCTCATATGAGCTAGACAGCCTTCACCCATGTCATGAGCATAGATTG 180
QY      181 TCACAGAGAGGTCATGAGAGCTGTTTATCTAGAACTTCATCTTGTGCGATTCAACAGAGT 240
DB      181 TCACAGAGAGGTCATGAGAGCTGTTTATCTAGAACTTCATCTTGTGCGATTCAACAGAGT 240
QY      241 GAAGACCATGAGGACCATCATGATGATGATGATGATGATGATGATGATGATGATGATG 300
DB      241 GAAGACCATGAGGACCATCATGATGATGATGATGATGATGATGATGATGATGATGATG 300
QY      301 GGTGTGATCTATTGGCGAAGCTGTACAAATGTCACCCCAAAAAAATGCTGATGCTGCT 360
DB      301 GGTGTGATCTATTGGCGAAGCTGTACAAATGTCACCCCAAAAAAATGCTGATGCTGCT 360
QY      361 GACCTGGGAGCTGTTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
DB      361 GACCTGGGAGCTGTTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
QY      421 TGGTGGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480
DB      421 TGGTGGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480
```

```
QY      481 ATTTTACAGAGGAGTGCACACCGATGCTGATATCCCGAGCAGTATGATGCTGTTGTT 540
DB      481 ATTTTACAGAGGAGTGCACACCGATGCTGATATCCCGAGCAGTATGATGCTGTTGTT 540
QY      541 ATTTGCTATGGAAGAGTGTGCGCCAGAAACAGAGAGTGTGCTATGATGATGATGATGATG 600
DB      541 ATTTGCTATGGAAGAGTGTGCGCCAGAAACAGAGAGTGTGCTATGATGATGATGATGATG 600
QY      601 GATTAACCTGAGCATTTATGCGCCATTTGACGTAGATTTATATACACCAACCAAGGCT 660
DB      601 GATTAACCTGAGCATTTATGCGCCATTTGACGTAGATTTATATACACCAACCAAGGCT 660
QY      661 CTTCATCATGATGCTGACAGGACCATGATTTCAACCAAGTGGTTATGATGATGATGATG 720
DB      661 CTTCATCATGATGCTGACAGGACCATGATTTCAACCAAGTGGTTATGATGATGATGATG 720
QY      721 GTGTCTACCATTTAGAGGATTTAATGTAAGTATGATGATGATGATGATGATGATGATG 780
DB      721 GTGTCTACCATTTAGAGGATTTAATGTAAGTATGATGATGATGATGATGATGATGATG 780
QY      781 GTACAGATCTACGCAACGAGTGCATGACAGTGTGATCTAGAGATATTTACATGATGATG 840
DB      781 GTACAGATCTACGCAACGAGTGCATGACAGTGTGATCTAGAGATATTTACATGATGATG 840
QY      841 GATCCTGATGAGATTTACACAGATGATGATGATGATGATGATGATGATGATGATGATG 900
DB      841 GATCCTGATGAGATTTACACAGATGATGATGATGATGATGATGATGATGATGATGATG 900
QY      901 CATGGCTGTGCAATGAAAGCCACAGATATTCCTGATCATATTCCTGATGATGATGATG 960
DB      901 CATGGCTGTGCAATGAAAGCCACAGATATTCCTGATCATATTCCTGATGATGATGATG 960
QY      961 GTCTGGGAGCTTCTATGTCATTCACCAAGAGACATCTTCAATTAAGTGGCAAAAT 1020
DB      961 GTCTGGGAGCTTCTATGTCATTCACCAAGAGACATCTTCAATTAAGTGGCAAAAT 1020
QY      1021 GCTGTGTGTGCTATGAGAGTGTGACAGAGCCATTCAGAAATTTGTCACAGATTAATGT 1080
DB      1021 GCTGTGTGTGCTATGAGAGTGTGACAGAGCCATTCAGAAATTTGTCACAGATTAATGT 1080
QY      1081 GTTGTGACAGCTTAAATACAGATTTACTGTTTACATGATGATGATGATGATGATGATG 1140
DB      1081 GTTGTGACAGCTTAAATACAGATTTACTGTTTACATGATGATGATGATGATGATGATG 1140
QY      1141 TTGCAACAGACCCCTGGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
DB      1141 TTGCAACAGACCCCTGGAGGATGATGATGATGATGATGATGATGATGATGATGATG 1200
QY      1201 AGTGAGAGCTTGTACATCATGCTGATGCTGATGCTGATGATGATGATGATGATGATG 1260
DB      1201 AGTGAGAGCTTGTACATCATGCTGATGCTGATGCTGATGATGATGATGATGATGATG 1260
QY      1261 CATGGCTGTGCTTCTTAGGACAGATGATGATGATGATGATGATGATGATGATGATGATG 1320
DB      1261 CATGGCTGTGCTTCTTAGGACAGATGATGATGATGATGATGATGATGATGATGATGATG 1320
QY      1321 AGACCTGGTACCTTGGTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1380
DB      1321 AGACCTGGTACCTTGGTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1380
QY      1381 ACCTTGTATCTTAAATGTTGAAGATCTTAAGCAGACCTTCTTGTGATTAATGATGATG 1440
DB      1381 ACCTTGTATCTTAAATGTTGAAGATCTTAAGCAGACCTTCTTGTGATTAATGATGATG 1440
QY      1441 GGCATGAGGTGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1491
DB      1441 GGCATGAGGTGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1491
```

RESULT 3
ABL06474
ID ABL06474 standard: cDNA: 3211 BP.

```
XX ABL06474;
AC 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 13904.
DE
XX Drosophila: developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX P-PSDB; ABB62371.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 13904; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
XX sequences (ABB5737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 3211 BP; 922 A; 708 C; 688 G; 893 T; 0 other:
SQ
XX
XX Query Match 1.7%; Score 26; DB 23; Length 3211;
XX Best Local Similarity 100.0%; Pred. No. 0.011;
XX Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 922 CACAGATATTCCTGCATCATATCTT 947
DB 2629 CACAGATATTCCTGCATCATATCTT 2654
XX
XX
XX RESULT 4
XX ABL10542
XX ID ABL10542 standard; cDNA; 6398 BP.
XX
XX ABL10542;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 26108.
XX
XX Drosophila: developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX
```

```
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX P-PSDB; ABB6439.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 26108; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
XX sequences (ABB5737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 6398 BP; 2021 A; 1289 C; 1522 G; 1566 T; 0 other:
SQ
XX
XX Query Match 1.7%; Score 26; DB 23; Length 6398;
XX Best Local Similarity 100.0%; Pred. No. 0.011;
XX Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 922 CACAGATATTCCTGCATCATATCTT 947
DB 665 CACAGATATTCCTGCATCATATCTT 690
XX
XX
XX RESULT 5
XX ABL23878
XX ID ABL23878 standard; DNA; 2171 BP.
XX
XX ABL23878;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 23107.
XX
XX Drosophila: developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX
```


PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148665.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149923.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150684.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151308.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.

PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161044.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.4%; Score 21; DB 21; Length 1599;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 77 ATGATGATGATGATGAACAA 97
|||||
DB 139 ATGATGATGATGATGAACAA 119

RESULT 7
ID ABL15963 standard; cDNA; 3309 BP.
XX ABL15963;
XX
XX 26-MAR-2002 (first entry)
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 42371.
XX
XX Drosophila: developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
OS Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001MO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PMD, Myers EW;
PI MPI; 2001-656860/75.
DR P-PSDB; ABB71860.
XX
XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
PT
XX
XX
PS Claim 1: SEQ ID NO 42371; 21bp + Sequence Listing: English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (AB57737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 3309 BP; 897 A; 893 C; 874 G; 645 T; 0 other;

Query Match 1.4%; Score 21; DB 23; Length 3309;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 37 GACGACCATGACGACGAC 57
|||||
DB 2674 GACGACCATGACGACGAC 2694

RESULT 8
AAT18924/c
ID AAT18924 standard; DNA; 4909 BP.
AC AAT18924;
XX 17-JAN-1997 (first entry)
XX
XX Plasmid pA1261.
XX
XX Spider; dragline protein; variant; monomer; polymer; circular;
XX fibre forming region; Spidroin 1; Nephila clavipes; Dp1; mimic;
XX DP-1A analogue; fibre; high tensile strength; elasticity; clothing;
XX rope; surgical suture; implant; reinforcement; film; coating; ss.
XX Synthetic.
XX
XX MO9429450-A2.
XX
XX 22-DEC-1994.
XX
XX 15-JUN-1994; 94WO-US06689.
XX
XX 15-JUN-1993; 93US-0077600.
XX
XX (DUPO) DU PONT DE NEMOURS & CO E I.
XX
XX Fahnestock SR;
XX
XX WPI: 1995-036479/05.
XX
XX
XX New synthetic variants of spider dragline protein - for making
XX fibres useful as clothing, surgical silk, plastic reinforcement
XX etc., also related DNA, vectors and transformed cells
XX
XX Example 1; Fig 13; 168bp; English.

This sequence represents the complete nucleotide sequence of the plasmid
CC pA1261. This plasmid was used in the construction of the vector pFP10
CC and DP-1B.9. pA1261 comprises a replication origin active in E. coli,
CC and DP-1B.9. pA1261 comprises a replication origin active in E. coli,
CC a selectable genetic marker which is a gene conferring resistance to
CC ampicillin, sites for the restriction endonucleases BamHI and BglII with
CC no essential sequences between them, and a third restriction site for
CC PstI, located within the selectable marker which produces cohesive ends
CC incompatible with those produced by BamHI and BglII. The polypeptide
CC monomers are variants based on a consensus sequence derived from the
CC fibre forming regions of spider dragline protein, esp. the natural
CC Protein 1 (Spidroin 1) from Nephila clavipes. Synthetic analogues
CC of DP1 were designed to mimic the repeating consensus sequence of the
CC natural protein and the pattern of variation among individual repeats.
CC DP-1A analogues are composed of a tandemly repeated 101 amino acid
CC monomer which comprises four repeats which differ from the consensus
CC sequence given in AAM06201, according to the pattern (1)-(5):
CC (1) the poly-alanine sequence varies in length from 0-7
CC residues; (2) when the entire poly-alanine sequence is deleted,
CC so also is the surrounding sequence encompassing AGGAGGAGGAGG;
CC (3) aside from the poly-alanine sequence, deletions usually
CC encompass integral multiples of three consecutive residues;
CC (4) deletion of GYG is generally accompanied by deletion of GRG
CC in the same sequence; and
CC (5) a repeat in which the entire poly-alanine sequence is
CC deleted is generally preceded by a repeat containing six alanine

CC residues.
CC The proteins may be used to produce fibres of high tensile strength and
CC elasticity, suitable for clothing, rope, surgical sutures, biomaterials
CC for implants, plastic reinforcements, films, coatings, etc.
XX

SO Sequence 4909 BP; 1216 A; 1251 C; 1258 G; 1164 T; 0 other;

Query Match 1.4%; Score 21; DB 16; Length 4909;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1332 CCTTGGTGTATATGATGA 1352
|||||
DB 2163 CCTTGGTGTATATGATGA 2143

RESULT 9
ABL15962/c
ID ABL15962 standard; cDNA; 6699 BP.
XX
XX ABL15962;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide seq ID NO 42368.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI: 2001-656860/75.
XX
XX P-PSDB; ABB71859.
XX
XX

New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 42368; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511); expressed DNA
CC sequences (ABB57737-ABB72072).
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/publshd_pcl_sequences.
XX
XX Sequence 6699 BP; 1663 A; 1551 C; 1541 G; 1944 T; 0 other;

Query Match 1.4%; Score 21; DB 23; Length 6699;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 37 GACGACCATGACGACGAC 57
|||||

DB 1636 GACCACCATGACGACCCAC 1616

RESULT 10

ABLS1954 standard; DNA; 40023 BP.

AC ABL51954;

DT 11-JUL-2002 (first entry)

XX Human solute carrier family 18 member 2 (SLC18A2) gene SEQ ID NO:1.

XX Human: solute carrier family 18 member 2; SLC18A2; vesicular monoamine;

XX vesicular monoamine transporter; VMAT2; polymorphic site; SNP;

XX single nucleotide polymorphism; anti-inflammatory; neuroleptic;

XX haplotyping; genotyping; respiratory inflammatory disease;

XX neuropsychiatric disorder; monoaminergic brain system; gene; ds.

XX Homo sapiens.

XX Key

XX misc_feature

XX misc_feature

XX misc_feature

XX misc_feature

XX misc_feature

XX misc_feature

XX misc_feature

XX misc_feature

XX misc_feature

XX misc_feature

XX misc_feature

XX misc_feature

XX misc_feature

XX misc_feature

XX misc_feature

XX misc_feature

XX misc_feature

XX misc_feature

XX misc_feature

XX misc_feature

XX misc_feature

XX misc_feature

XX misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

29050 /tag- r

29230 /note- "PS18: polymorphic base G or T"

29230 /tag- s

29395 /note- "PS19: polymorphic base C or T"

29395 /tag- t

32343 /note- "PS20: polymorphic base C or T"

32572 /tag- u

39464 /note- "PS21: polymorphic base C or T"

39766 /tag- v

39848 /note- "PS22: polymorphic base A or G"

39910 /tag- w

39984 /note- "PS23: polymorphic base T or C"

39984 /tag- x

39984 /note- "PS24: polymorphic base C or T"

39984 /tag- y

39984 /note- "PS25: polymorphic base A or C"

39984 /tag- z

39984 /note- "PS26: polymorphic base T or C"

39984 /tag- aa

39984 /note- "PS27: polymorphic base T or C"

39984 /tag- ab

39984 /note- "PS28: polymorphic base C or T"

39984 /tag- ac

39984 /note- "PS29: polymorphic base G or T"

39984 /tag- ad

39984 /note- "PS30: polymorphic base C or T"

39984 /tag- ae

39984 /note- "PS31: polymorphic base T or C"

39984 /tag- af

39984 /note- "PS32: polymorphic base C or T"

39984 /tag- ag

39984 /note- "PS33: polymorphic base T or C"

39984 /tag- ah

39984 /note- "PS34: polymorphic base A or C"

39984 /tag- ai

39984 /note- "PS35: polymorphic base G or C"

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

```

Query Match      1.43; Score 21; DB 24; Length 40023;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 GACGACCATGATGATGAT 90
    |||
Db 19958 GACGACCATGATGATGAT 19978

RESULT 11
AAH45894/C
ID AAH45894 standard; DNA; 65 BP.
XX
XX AAH45894;
AC
XX
XX 17-SEP-2001 (first entry)
DE
XX C parvum coding sequence PCR primer JCA302.
XX
XX Cryptosporidium parvum infection; enteric pathogen; vaccine; cattle;
KM epitope; PCR primer; ss.
XX
XX Cryptosporidium parvum.
OS
XX WO200145735-A2.
XX
XX 28-JUN-2001.
PD
XX
XX 20-DEC-2000; 2000WO-EP13387.
PF
XX
XX 21-DEC-1999; 99US-0171399.
PR
XX
XX (MERI-) MERIAL.
XX
XX Audonnet J, Gallo G;
PI
XX WPI: 2001-417978/44.
DR
XX
XX New immunological or vaccine composition for prevention, treatment and
PT control of Cryptosporidium parvum and/or enteric infections in bovines,
PT canines, has antigens from Cryptosporidium and another enteric pathogen
PT
XX
XX Example 3; Page 34; 66pp; English.
XX
XX The present invention relates to a vaccine against Cryptosporidium
CC parvum, which causes enteric infections in cattle, comprising two
CC epitopes or antigens from the microorganism. This can also be used to
CC treat enteric infections in canines, felines and equines. The present
CC sequence is a PCR primer used to obtain the coding sequence of an epitope
CC useful in the composition of the invention.
XX
XX Sequence 65 BP; 9 A; 14 C; 14 G; 28 T; 0 other;
SQ
Query Match      1.38; Score 20; DB 22; Length 65;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 GACCATGATGATGATGATGA 92
    |||
Db 42 GACCATGATGATGATGATGA 23

RESULT 12
ABL93659
ID ABL93659 standard; cDNA; 461 BP.
XX
XX ABL93659;
AC
XX
XX 10-JUN-2002 (first entry)
DT
XX
XX Arabidopsis thaliana nucleic acid sequence Ref:2027424 SEQ ID NO:424.
DE

```

```

XX Arabidopsis thaliana; insecticide; fungicide; plant; mapping; diagnosis;
KM genetic modification; gene; ss.
XX Arabidopsis thaliana.
OS
XX US2002023280-A1.
XX
XX 21-FEB-2002.
PD
XX
XX 26-JAN-2001; 2001US-0770444.
PF
XX
XX 27-JAN-2000; 2000US-178502P.
PR
XX
XX (GORL/) GORLACH J.
PA (ANY/) AN Y.
PA (HAM/) HAMILTON C M.
PA (PRIC/) PRICE J L.
PA (RAIN/) RAINES T M.
PA (YUY/) YU Y.
PA (RAME/) RAMFARA J G.
PA (PAGE/) PAGE A.
PA (MATH/) MATHER A V.
PA (LEDE/) LEDFORD B L.
PA (WOES/) WOESSNER J P.
PA (HAAS/) HAAS W D.
PA (GARC/) GARCIA C A.
PA (KRIC/) KRICKER M.
PA (SLAT/) SLATER T.
PA (DAVI/) DAVIS K R.
PA (ALEE/) ALLEN K.
PA (HOEF/) HOFFMAN N.
PA (HURB/) HURBAN P.
XX
XX Goriach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
PI Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
PI Hurban P;
XX
XX WPI: 2002-267486/31.
DR
XX
XX New Arabidopsis thaliana nucleic acid, for identifying homologous
PT genes, producing compositions that modulate the expression or function
PT of its encoded protein, and mapping functional regions of a protein -
PT
XX
XX Claim 1; SEQ ID 424; 44pp; English.
PS
XX
XX The present invention describes an Arabidopsis thaliana nucleic acid (I)
CC comprising a sequence capable of hybridising under stringent conditions
CC to a sequence (SI) selected from any one of the 999 sequences given in
CC AB183236 to AB194234 (I) have insecticide and fungicide activities, and
CC they can be used as protein expression modulators. (I) can be used in
CC identifying homologous or related genes, in producing compositions that
CC modulate the expression or function of their encoded proteins, mapping
CC functional regions of the proteins, and in studying associated
CC physiological pathways. (I) can also be used: (1) for the genetic
CC manipulation of cells, particularly plant cells; (2) in screening assays
CC of various plant strains to determine the strains that are best capable
CC of withstanding a particular disease or environmental stress; (3) for
CC enhancing or inhibiting production of a biosynthetic product in a plant;
CC (4) as probes in mapping and in diagnosis, in genetic modification and
CC for screening purposes, to generate additional copies of the nucleic
CC acids, to generate ribozymes or antisense oligonucleotides, and as
CC single-stranded DNA probes or as triple-strand forming oligonucleotides;
CC and (5) for generating genetically modified transgenic organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
XX Sequence 461 BP; 147 A; 77 C; 122 G; 115 T; 0 other;
SQ
Query Match      1.34; Score 20; DB 24; Length 461;
Best Local Similarity 100.0%; Pred. No. 12;

```


PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157723.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160757.
PR 21-OCT-1999; 99US-0160758.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160960.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.3%; Score 20; DB 21; Length 706;
Best Local Similarity 100.0%; Pred No. 12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 77 ATATGATGATGATGAACA 96
Db 679 ATGATGATGATGATGAACA 698

RESULT 14
AAC34259
ID AAC34259 standard; DNA; 708 BP.
XX
AC AAC34259;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 6012.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126264.
PR 01-APR-1999; 99US-0126785.
PR 06-APR-1999; 99US-0127462.
PR 08-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137529.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.


```
XX DNA encoding novel human diagnostic protein #3301.
DE
XX
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX
XX WPI; 2001-639362/73.
XX
XX P-PSDB; ABG03310.
XX
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic; gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX
XX Claim 1; SEQ ID NO 3301; 103pp; English.
XX
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensic; gene mapping, identification of mutations
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197/AAS94364 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 879 BP; 227 A; 196 C; 225 G; 231 T; 0 other;
XX
XX
XX Query Match 1.3%; Score 20; DB 23; Length 879;
XX Best Local Similarity 100.0%; Pred. No. 12;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX
XX 77 ATGATGATGATGATGAACA 96
XX ||||||||||||||||||
XX Db 655 ATGATGATGATGATGAACA 674
```

Search completed: November 29, 2002, 10:09:02
Job time : 368 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2002, 08:54:48 (Search time 60 Seconds
(without alignments)
7620.918 Million cell updates/sec

Title: US-09-868-760-6

Sequence: 1 gctgtgttataagagagcgca 1491

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 441362 seqs, 15338381 residues

Word size: 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: Issued Patents NA: *
1: /cgn2.6/prodata/1/lna/5A.COMB.seq: *
2: /cgn2.6/prodata/1/lna/5B.COMB.seq: *
3: /cgn2.6/prodata/1/lna/6A.COMB.seq: *
4: /cgn2.6/prodata/1/lna/6B.COMB.seq: *
5: /cgn2.6/prodata/1/lna/PCNUS.COMB.seq: *
6: /cgn2.6/prodata/1/lna/Backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	1.4	4909	4 US-08-556-978B-78	Sequence 78, Appl
2	19	1.3	53	2 US-08-687-865A-10	Sequence 10, Appl
3	19	1.3	53	4 US-09-043-711-10	Sequence 10, Appl
4	19	1.3	55	2 US-08-687-865A-11	Sequence 11, Appl
5	19	1.3	55	4 US-09-043-711-11	Sequence 11, Appl
6	19	1.3	96	3 US-08-737-336-2	Sequence 2, Appl
7	19	1.3	105	2 US-08-687-865A-23	Sequence 23, Appl
8	19	1.3	105	4 US-09-043-711-23	Sequence 23, Appl
9	19	1.3	129	2 US-08-679-865-38	Sequence 38, Appl
10	19	1.3	129	4 US-09-263-975-38	Sequence 38, Appl
11	19	1.3	133	2 US-08-680-876-41	Sequence 41, Appl
12	19	1.3	133	4 US-08-680-876-41	Sequence 41, Appl
13	19	1.3	153	2 US-08-687-865A-20	Sequence 20, Appl
14	19	1.3	234	4 US-09-043-711-20	Sequence 20, Appl
15	19	1.3	234	2 US-08-687-865A-20	Sequence 20, Appl
16	19	1.3	466	1 US-08-457-245-20	Sequence 20, Appl
17	19	1.3	782	4 US-09-004-723-51	Sequence 51, Appl
18	19	1.3	782	2 US-08-749-699-51	Sequence 91, Appl
19	19	1.3	782	4 US-09-004-723-51	Sequence 91, Appl
20	19	1.3	782	2 US-08-749-699-51	Sequence 91, Appl
21	19	1.3	783	4 US-09-004-723-53	Sequence 93, Appl
22	19	1.3	783	2 US-08-749-699-53	Sequence 93, Appl
23	19	1.3	783	4 US-09-004-723-93	Sequence 93, Appl
24	19	1.3	811	4 US-09-004-731-88	Sequence 88, Appl
25	19	1.3	811	2 US-08-749-699-88	Sequence 88, Appl
26	19	1.3	811	4 US-09-004-723-88	Sequence 88, Appl
27	19	1.3	833	4 US-09-004-731-90	Sequence 90, Appl

28	19	1.3	823	4 US-08-749-699-90	Sequence 90, Appl
29	19	1.3	823	2 US-09-004-729-90	Sequence 90, Appl
30	19	1.3	1056	2 US-08-701-191A-5	Sequence 25, Appl
31	19	1.3	1395	1 US-07-991-867B-25	Sequence 25, Appl
32	19	1.3	1395	1 US-08-107-755A-25	Sequence 25, Appl
33	19	1.3	1395	2 US-08-544-332-25	Sequence 25, Appl
34	19	1.3	1395	4 US-09-370-861A-25	Sequence 28, Appl
35	19	1.3	1478	4 US-09-545-814-28	Sequence 30, Appl
36	19	1.3	1478	4 US-09-545-814-30	Sequence 30, Appl
37	19	1.3	1599	3 US-08-737-336-5	Sequence 7, Appl
38	19	1.3	1896	1 US-08-605-541B-11	Sequence 11, Appl
39	19	1.3	2081	2 US-09-096-982-7	Sequence 7, Appl
40	19	1.3	2081	2 US-08-653-650A-7	Sequence 7, Appl
41	19	1.3	4770	4 US-09-000-09A-45	Sequence 45, Appl
42	19	1.3	6768	1 US-08-107-755A-1	Sequence 1, Appl
43	19	1.3	8457	1 US-07-991-867B-1	Sequence 1, Appl
44	19	1.3	8457	2 US-08-544-332-1	Sequence 1, Appl
45	19	1.3	8457	4 US-09-370-861A-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-556-978B-78/c
Sequence 78, Application US/08556978B
Patent No. 6268169
GENERAL INFORMATION:
APPLICANT: FAHNESTOCK, STEPHEN F.
TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED
NUMBER OF INVENTIONS: SPIDER SILK ANALOGS
TITLE OF INVENTION: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: UNITED STATES OF AMERICA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WINDOWS 95
SOFTWARE: MICROSOFT WORD FOR WINDOWS 95
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/556,978B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/077,600
FILING DATE: JUNE 15, 1993
ATTORNEY/AGENT INFORMATION:
NAME: FLOYD, LINDA AXAMETHY
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CR-9389-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 78:
SEQUENCE CHARACTERISTICS:
LENGTH: 4909 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
US-08-556-978B-78
Query Match 1.4%: Score 21; DB 4; Length 4909;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1332 CCTGTGTATGTTATAGATGA 1352
|||||

Db 2163 CCTTGCTGATGTTATGATGA 2143

RESULT 2

US-08-687-865A-10/C

Sequence 10, Application US/08687865A

Patent No. 5955596

GENERAL INFORMATION:

APPLICANT: Jones, Kevin F.

APPLICANT: Zagursky, Robert J.

TITLE OF INVENTION: The Nuca Protein of Haemophilus

TITLE OF INVENTION: Influenzae and the Gene Encoding That Protein

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: American Cyanamid Company

STREET: One Cyanamid Plaza

CITY: Wayne

STATE: New Jersey

COUNTRY: U.S.A.

ZIP: 07470

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/687,865A

FILING DATE: 26-JUL-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Gordon, Alan M.

REGISTRATION NUMBER: 30,637

REFERENCE/DOCKET NUMBER: 33,250-00

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-831-3244

TELEFAX: 201-831-3305

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 53 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-687-865A-10

Query Match

Best Local Similarity 1.3%; Score 19; DB 2; Length 53;

Best Local Similarity 100.0%; Pred. No. 6.1;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ACCATGATGATGATGATGA 92

DB 49 ACCATGATGATGATGATGA 31

RESULT 3

US-09-043-711-10/C

Sequence 10, Application US/09043711

Patent No. 6221365

GENERAL INFORMATION:

APPLICANT: Jones, Kevin F.

APPLICANT: Zagursky, Robert J.

TITLE OF INVENTION: The Nuca Protein of Haemophilus

TITLE OF INVENTION: Influenzae and the Gene Encoding That Protein

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: American Cyanamid Company

STREET: One Cyanamid Plaza

CITY: Wayne

STATE: New Jersey

COUNTRY: U.S.A.

ZIP: 07470

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/043,711

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/687,865

FILING DATE: 26-JUL-1996

ATTORNEY/AGENT INFORMATION:

NAME: Gordon, Alan M.

REGISTRATION NUMBER: 30,637

REFERENCE/DOCKET NUMBER: 33,250-00

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-831-3244

TELEFAX: 201-831-3305

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 53 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-09-043-711-10

Query Match

Best Local Similarity 1.3%; Score 19; DB 4; Length 53;

Best Local Similarity 100.0%; Pred. No. 6.1;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ACCATGATGATGATGATGA 92

DB 49 ACCATGATGATGATGATGA 31

RESULT 4

US-08-687-865A-11

Sequence 11, Application US/08687865A

Patent No. 5955596

GENERAL INFORMATION:

APPLICANT: Jones, Kevin F.

APPLICANT: Zagursky, Robert J.

TITLE OF INVENTION: The Nuca Protein of Haemophilus

TITLE OF INVENTION: Influenzae and the Gene Encoding That Protein

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: American Cyanamid Company

STREET: One Cyanamid Plaza

CITY: Wayne

STATE: New Jersey

COUNTRY: U.S.A.

ZIP: 07470

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/687,865A

FILING DATE: 26-JUL-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Gordon, Alan M.

REGISTRATION NUMBER: 30,637

REFERENCE/DOCKET NUMBER: 33,250-00

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-831-3244

TELEFAX: 201-831-3305

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 55 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-687-865A-11

Query Match 1.3%; Score 19; DB 2; Length 55;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 ACCATGATGATGATGATGA 92
Db 9 ACCATGATGATGATGATGA 27

RESULT 5
US-09-043-711-11
Sequence 11, Application US/09043711
Patent No. 6221965

GENERAL INFORMATION:
APPLICANT: Jones, Kevin F.
APPLICANT: Zagursky, Robert J.
TITLE OF INVENTION: The Nuca Protein of Haemophilus
TITLE OF INVENTION: Influenzae and the Gene Encoding That Protein
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: American Cyanamid Company
STREET: One Cyanamid Plaza
CITY: Wayne
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07470

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/043,711
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/687,865
FILING DATE: 26-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gordon, Alan M.
REGISTRATION NUMBER: 30,637
REFERENCE/DOCKET NUMBER: 33,250-00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-831-3305
TELEFAX: 201-831-3244

INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 55 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-043-711-11

Query Match 1.3%; Score 19; DB 4; Length 55;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 ACCATGATGATGATGATGA 92
Db 9 ACCATGATGATGATGATGA 27

RESULT 6
US-08-737-336-2/C

Sequence 2, Application US/08737336
Patent No. 6013262
GENERAL INFORMATION:
APPLICANT: FRAZER, Ian
APPLICANT: ZHOU, Jian
TITLE OF INVENTION: RECOMBINANT PAPILLOMA VIRUS L1
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/737,336
FILING DATE: 16-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/AU95/00292
FILING DATE: 17-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PM 5667
FILING DATE: 17-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Saxe, Bernhard D.
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 65064/118
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 96 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-737-336-2

Query Match 1.3%; Score 19; DB 3; Length 96;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 ACCATGATGATGATGATGA 92
Db 33 ACCATGATGATGATGATGA 15

RESULT 7
US-08-687-865A-23/C
Sequence 23, Application US/08687865A
Patent No. 5955596
GENERAL INFORMATION:
APPLICANT: Jones, Kevin F.
APPLICANT: Zagursky, Robert J.

TITLE OF INVENTION: The Nuca Protein of Haemophilus
TITLE OF INVENTION: Influenzae and the Gene Encoding That Protein
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: American Cyanamid Company
STREET: One Cyanamid Plaza
CITY: Wayne
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07470

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,865A
FILING DATE: 26-JUL-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Gordon, Alan M.
REGISTRATION NUMBER: 30,637
REFERENCE/DOCKET NUMBER: 33,250-00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-831-3244
TELEFAX: 201-831-3305
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 105 base pairs
STRANDEDNESS: single
TYPE: nucleic acid
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-687-865A-23

Query Match 1.3%; Score 19; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 74 ACCATGATGATGATGA 92
Db 36 ACCATGATGATGATGA 18

RESULT 8
US-09-043-711-23/C
Sequence 23, Application US/09043711
Patent No. 6221365
GENERAL INFORMATION:
APPLICANT: Jones, Kevin F.
APPLICANT: Zagursky, Robert J.
TITLE OF INVENTION: The NUCA Protein of Haemophilus
TITLE OF INVENTION: Influenzae and the Gene Encoding That Protein
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: American Cyanamid Company
STREET: One Cyanamid Plaza
CITY: Wayne
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07470
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/043,711
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/687,865
FILING DATE: 26-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gordon, Alan M.
REGISTRATION NUMBER: 30,637
REFERENCE/DOCKET NUMBER: 33,250-00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-831-3244
TELEFAX: 201-831-3305
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 105 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-043-711-23

Query Match 1.3%; Score 19; DB 4; Length 105;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 74 ACCATGATGATGATGA 92
Db 36 ACCATGATGATGATGA 18

RESULT 9
US-08-679-865-38/C
Sequence 38, Application US/08679865
Patent No. 5912137
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
APPLICANT: Cubitt, Andrew B.
TITLE OF INVENTION: Assays for Protein Kinases Using
TITLE OF INVENTION: Fluorescent Protein Substrates
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/679,865
FILING DATE: 16-JUL-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Storella, John S.
REGISTRATION NUMBER: 32,944
REFERENCE/DOCKET NUMBER: 023072-069000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 129 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 13...129
US-08-679-865-38

Query Match 1.3%; Score 19; DB 2; Length 129;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 74 ACCATGATGATGATGA 92
Db 45 ACCATGATGATGATGA 27

RESULT 10
US-08-680-876-38/C
Sequence 38, Application US/08680876
Patent No. 5925558

```

: GENERAL INFORMATION:
: APPLICANT: Tsien, Roger Y.
: APPLICANT: Cubitt, Andrew B.
: TITLE OF INVENTION: Assays for Protein Kinases Using
: TITLE OF INVENTION: Fluorescent Protein Substrates
: NUMBER OF SEQUENCES: 48
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/680,876
: FILING DATE: 16-JUL-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Storella, John S.
: REGISTRATION NUMBER: 32,944
: REFERENCE/DOCKET NUMBER: 023072-069200
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 38:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 129 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 13..129
: US-08-680-876-38

Query Match      1.3%; Score 19; DB 2; Length 129;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 74 ACCATGATGATGATGATGA 92
Db 45 ACCATGATGATGATGATGA 27

RESULT 11
US-09-263-975-38/C
: Sequence 38, Application US/09263975
: Patent No. 6248550
: GENERAL INFORMATION:
: APPLICANT: Tsien, Roger Y.
: APPLICANT: Cubitt, Andrew B.
: TITLE OF INVENTION: Assays for Protein Kinases Using
: TITLE OF INVENTION: Fluorescent Protein Substrates
: NUMBER OF SEQUENCES: 48
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
```

```

: APPLICATION NUMBER: US/09/263,975
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/679,865
: FILING DATE: 16-JUL-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Storella, John S.
: REGISTRATION NUMBER: 32,944
: REFERENCE/DOCKET NUMBER: 023072-069000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0300
: TELEFAX: (415) 576-0200
: INFORMATION FOR SEQ ID NO: 38:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 129 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 13..129
: US-09-263-975-38

Query Match      1.3%; Score 19; DB 4; Length 129;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 74 ACCATGATGATGATGATGA 92
Db 45 ACCATGATGATGATGATGA 27

RESULT 12
US-08-679-865-41/C
: Sequence 41, Application US/08679865
: Patent No. 5912137
: GENERAL INFORMATION:
: APPLICANT: Tsien, Roger Y.
: APPLICANT: Cubitt, Andrew B.
: TITLE OF INVENTION: Assays for Protein Kinases Using
: TITLE OF INVENTION: Fluorescent Protein Substrates
: NUMBER OF SEQUENCES: 48
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/679,865
: FILING DATE: 16-JUL-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Storella, John S.
: REGISTRATION NUMBER: 32,944
: REFERENCE/DOCKET NUMBER: 023072-069000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 41:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 153 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
```

Mon Dec 2 12:19:18 2002

us-09-868-760-6.oli.rni

Page 6

```
; MOLECULE TYPE: DNA
; US-08-679-865-41
Query Match 1.3%; Score 19; DB 2; Length 153;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ACCATGATGATGATGATGA 92
   |||
Db 45 ACCATGATGATGATGATGA 27

RESULT 13
US-08-680-876-41/C
; Sequence 41, Application US/08680876
; Patent No. 5925558
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Cubitt, Andrew B.
; TITLE OF INVENTION: Assays for Protein Kinases Using
; TITLE OF INVENTION: Fluorescent Protein Substrates
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680,876
; FILING DATE: 16-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Storella, John S.
; REGISTRATION NUMBER: 32,944
; REFERENCE/DOCKET NUMBER: 023072-069200
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 153 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-680-876-41

Query Match 1.3%; Score 19; DB 2; Length 153;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ACCATGATGATGATGATGA 92
   |||
Db 45 ACCATGATGATGATGATGA 27

RESULT 14
US-09-263-975-41/C
; Sequence 41, Application US/09263975
; Patent No. 6248550
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Cubitt, Andrew B.
; TITLE OF INVENTION: Assays for Protein Kinases Using
; TITLE OF INVENTION: Fluorescent Protein Substrates
; NUMBER OF SEQUENCES: 48
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/263,975
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/679,865
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Storella, John S.
; REGISTRATION NUMBER: 32,944
; REFERENCE/DOCKET NUMBER: 023072-069000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 153 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-09-263-975-41

Query Match 1.3%; Score 19; DB 4; Length 153;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ACCATGATGATGATGATGA 92
   |||
Db 45 ACCATGATGATGATGATGA 27

RESULT 15
US-08-687-865A-20/C
; Sequence 20, Application US/08687865A
; Patent No. 5935596
; GENERAL INFORMATION:
; APPLICANT: Jones, Kevin F.
; APPLICANT: Zagursky, Robert J.
; APPLICANT: Ooi, Peggy
; TITLE OF INVENTION: The Nuca Protein of Haemophilus
; TITLE OF INVENTION: Influenzae and the Gene Encoding That Protein
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07470
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,865A
; FILING DATE: 26-JUL-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Gordon, Alan M.
```

Mon Dec 2 12:19:18 2002

us-09-868-760-6.oli.rni

Page 7

REGISTRATION NUMBER: 30,637
REFERENCE/DOCKET NUMBER: 33,250-00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-831-3244
TELEFAX: 201-831-3305
INFORMATION FOR SRD ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 234 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 21..227
US-08-687-865A-20

Query Match 1.3%; Score 19; DB 2; Length 234;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 74 ACCATGATGATGATGATGA 92
|||||
Db 53 ACCATGATGATGATGATGA 35

Search completed: November 29, 2002, 10:08:57
Job time : 72 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2002, 09:09:37 : Search time 67 Seconds

(without alignments)
8570.256 Million cell updates/sec

Title: US-09-868-760-6

Perfect score: 1491
Sequence: 1 gaggggagcagtgtaacga.....gctgtgtacagcggagca 1491

Scoring table: OLIGO-MNC
Gapop 60.0 , Gapext 60.0

Searched: 341543 seqs, 192557720 residues

Word size : 0

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications, NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCCT_NEM_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEM_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEM_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEM_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEM_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEM_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEM_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	20	1.3	461	US-09-770-444-424
2	20	1.3	1356	US-09-938-842A-1266
3	19	1.3	104	US-09-864-761-23649
4	19	1.3	104	US-09-864-761-23649
5	19	1.3	129	US-09-864-761-28873
6	19	1.3	142	US-09-864-761-28873
7	19	1.3	133	US-09-864-761-28873
8	19	1.3	439	US-09-770-444-955
9	19	1.3	535	US-09-864-761-5917
10	19	1.3	535	US-09-864-761-5917
11	19	1.3	574	US-09-864-761-12293
12	19	1.3	1257	US-09-808-483-17
13	19	1.3	2151	US-09-845-157-1
14	19	1.3	3157	US-09-965-703-74
15	18	1.2	129	US-09-864-761-23649
16	18	1.2	129	US-09-864-761-23649
17	18	1.2	247	US-09-878-574-13298
18	18	1.2	257	US-09-764-869-399
19	18	1.2	262	US-09-878-574-15461

c 20	18	1.2	266	10	US-09-878-574-14896	Sequence 14896, A
c 21	18	1.2	286	10	US-09-878-574-5372	Sequence 5372, Ap
c 22	18	1.2	451	10	US-09-770-444-667	Sequence 667, App
c 23	18	1.2	453	10	US-09-924-035A-231	Sequence 231, App
c 24	18	1.2	507	10	US-09-867-701-988	Sequence 988, App
c 25	18	1.2	511	10	US-09-864-761-8859	Sequence 8859, Ap
c 26	18	1.2	547	10	US-09-864-761-8835	Sequence 8835, Ap
c 27	18	1.2	708	9	US-09-938-842A-1875	Sequence 1875, Ap
c 28	18	1.2	3072	10	US-09-841-739-3	Sequence 3, App11
c 29	18	1.2	3133	10	US-09-841-739-6	Sequence 6, App11
c 30	18	1.2	3612	10	US-09-841-739-4	Sequence 4, App11
c 31	18	1.2	3613	10	US-09-841-739-12	Sequence 12, App1
c 32	18	1.2	6301	9	US-10-029-413A-3	Sequence 3, App11
c 33	18	1.2	7193	9	US-10-029-413A-17	Sequence 17, App1
c 34	18	1.2	7193	9	US-10-029-413A-15	Sequence 15, App1
c 35	18	1.2	7362	9	US-10-029-413A-15	Sequence 3, App11
c 36	18	1.2	7560	12	US-10-100-912-3	Sequence 20, App1
c 37	18	1.2	7746	10	US-09-778-927A-20	Sequence 25, App1
c 38	18	1.2	32	10	US-09-995-297-25	Sequence 9, App11
c 39	17	1.1	57	10	US-09-905-114-9	Sequence 682, App
c 40	17	1.1	90	10	US-09-969-373-682	Sequence 27845, A
c 41	17	1.1	146	10	US-09-864-761-27845	Sequence 23373, A
c 42	17	1.1	151	10	US-09-864-761-23373	Sequence 24663, A
c 43	17	1.1	164	10	US-09-864-761-24663	Sequence 28856, A
c 44	17	1.1	198	10	US-09-864-761-28856	
c 45	17	1.1				

ALIGNMENTS

RESULT 1
US-09-770-444-424
Sequence 424, Application US/09770444
Patent No. US2002023280A1
GENERAL INFORMATION:
APPLICANT: Goriach, Joren
APPLICANT: Han, Yong-Qiang
APPLICANT: Hamilton, Carol M.
APPLICANT: Price, Jennifer L.
APPLICANT: Raines, Tracy M.
APPLICANT: Yu, Yang
APPLICANT: Rameaka, Joshua G.
APPLICANT: Page, Amy
APPLICANT: Mathew, Abraham V.
APPLICANT: Ledford, Brooke L.
APPLICANT: Woessner, Jeffrey P.
APPLICANT: Haas, William David
APPLICANT: Garcia, Carlos A.
APPLICANT: Kricker, Maia
APPLICANT: Slader, Ted
APPLICANT: Davis, Keith R.
APPLICANT: Allen, Keith
APPLICANT: Hoffman, Neil
APPLICANT: Hildan, Patrick
TITLE OR INVENTION: Expressed Sequences of Arabidopsis
FILE REFERENCE: 2027 (PARA-016PRV)
CURRENT APPLICATION NUMBER: US/09/770,444
PRIORITY FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: 60/178,502
NUMBER OF SEQ ID NOS: 999
SOFTWARE: FASTSD for Windows Version 4.0
SEQ ID NO 424
LENGTH: 461
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-770-444-424
Query Match 1.3%; Score 20; DB 10; Length 461;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 73 GACCATGATGATGATGCA 92
|||||
DB 322 GACCATGATGATGATGCA 341

RESULT 2
US-09-938-842A-1266/C
Sequence 1266, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938, 842A
PRIOR FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 1266
LENGTH: 1356
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-1266

Query Match 1.3%; Score 20; DB 9; Length 1356;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 77 ATGATGATGATGATGAACA 96
|||||
DB 259 ATGATGATGATGATGAACA 240

RESULT 3
US-09-864-761-23649/C
Sequence 23649, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 23649
LENGTH: 104
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC012619.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.5
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.8
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.3
OTHER INFORMATION: NT HIT: AF148532.1, EVALU8 3.50e+00
OTHER INFORMATION: EST_HUMAN HIT: BE963379.2, EVALU8 1.30e+00
US-09-864-761-23649

Query Match 1.3%; Score 19; DB 10; Length 104;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 ACCATGATGATGATGCA 92
|||||
DB 74 ACCATGATGATGATGCA 56

RESULT 4
US-09-864-761-24379/C
Sequence 24379, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30

```

; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24379
; LENGTH: 104
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC020919.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.7
; OTHER INFORMATION: NT HIT: AF148532.1, EVALUATE 3.50e+00
; OTHER INFORMATION: EST_HUMAN HIT: BE963379.2, EVALUATE 1.30e+00
; US-09-864-761-24379

Query Match      1.3%  Score 19; DB 10; Length 104;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      74 ACCATGATGATGATGATGA 92
      |||||||||||||||
DB      74 ACCATGATGATGATGATGA 56

RESULT 5
US-09-864-681-38/C
; Sequence 38 Application US/09884681
; Patent No. US20020061546A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; TITLE OF INVENTION: Assays for Protein Kinases Using
;             Fluorescent Protein Substrates
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/884,681
; FILING DATE: 19-Jun-2001
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/679,865
; FILING DATE: <unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Storella, John S.
; REGISTRATION NUMBER: 32,944
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 129 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 13..129
; US-09-864-681-38

Sequence Description: SEQ ID NO: 38:

Query Match      1.3%  Score 19; DB 10; Length 129;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      74 ACCATGATGATGATGATGA 92
      |||||||||||||||
DB      45 ACCATGATGATGATGATGA 27

RESULT 6
US-09-864-761-28873/C
; Sequence 28873, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;             GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-28
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
```

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2001-01-29
PRIOR APPLICATION NUMBER: US 09/774,203
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonex Sequence Listing Engine vers. 1.1
SEQ ID NO: 28873
LENGTH: 142
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC025644.2
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.84
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.59
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.76
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.57
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.57
US-09-864-761-28873

Query Match 1.3%; Score 19; DB 10; Length 142;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 ATGATGATGATGATGAAAC 95
DB 105 ATGATGATGATGATGAAAC 87

RESULT 7
US-09-884-681-41/C
Sequence 41, Application US/09884681
Patent No. US20020061546A1
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
Cubitt, Andrew B.
TITLE OF INVENTION: Assays for Protein Kinases Using
Fluorescent Protein Substrates
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/884,681
FILING DATE: 19-Jun-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/679,865
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Scirella, John S.
REGISTRATION NUMBER: 32,944
REFERENCE/DOCKET NUMBER: 023072-069000
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 153 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-09-884-681-41

Query Match 1.3%; Score 19; DB 10; Length 153;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ACCATGATGATGATGATGA 92
DB 45 ACCATGATGATGATGATGA 27

RESULT 8
US-09-770-444-955
Sequence 955, Application US/09770444
Patent No. US20020023280A1
GENERAL INFORMATION:
APPLICANT: Corlach, Jörn
APPLICANT: An, Yong-Olang
APPLICANT: Hamilton, Carol M.
APPLICANT: Price, Jennifer L.
APPLICANT: Rames, Tracy M.
APPLICANT: Yu, Tang
APPLICANT: Kameoka, Joshua G.
APPLICANT: Page, Amy
APPLICANT: Mathew, Abraham V.
APPLICANT: Ledford, Brooke L.
APPLICANT: Woessner, Jeffrey P.
APPLICANT: Haas, William David
APPLICANT: Garcia, Carlos A.
APPLICANT: Krickler, Maya
APPLICANT: Slader, Ted
APPLICANT: Davis, Keith R.
APPLICANT: Allen, Keith
APPLICANT: Hoffman, Neil
APPLICANT: Hurban, Patrick
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
FILE REFERENCE: 2027 (PARA-016PRV)
CURRENT APPLICATION NUMBER: US/09/770,444
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: 60/178,502
PRIOR FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 999
SOFTWARE: PasteSeq for Windows Version 4.0
SEQ ID NO: 955
LENGTH: 439
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(439)
OTHER INFORMATION: n = A,T,C or G
US-09-770-444-955

Query Match 1.3%; Score 19; DB 10; Length 439;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 837 ATATGATCCTCATGATAGAT 855
DB 373 ATATGATCCTCATGATAGAT 391

```
RESULT 9
US-09-864-761-6917/C
; Sequence 6917, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmca-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 6917
; LENGTH: 535
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC012619.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.3
US-09-864-761-6917

Query Match 1.3%; Score 19; DB 10; Length 535;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Db 412 ACCATGATGATGATGATCA 394

RESULT 10
US-09-864-761-7675/C
; Sequence 7675, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecmca-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7675
; LENGTH: 535
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC020919.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.7
US-09-864-761-7675

Query Match 1.3%; Score 19; DB 10; Length 535;
Best Local Similarity 100.0%; Pred. No. 5.3;
```

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ACCATGATGATGATGATGA 92
|||||
Db 412 ACCATGATGATGATGATGA 394

RESULT 11
US-09-864-761-12293/c
; Sequence 12293, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT FILING DATE: 2001-03-23
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 12293
; LENGTH: 574
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC025644.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.84
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.59
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.76
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.57
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.57
US-09-864-761-12293

Query Match 1.3%; Score 19; DB 10; Length 574;

Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 ATGATGATGATGATGAAC 95
|||||
Db 198 ATGATGATGATGATGAAC 180

RESULT 12
US-09-808-483-17
; Sequence 17, Application us/09808483
; Patent No. US2002001824A1
; GENERAL INFORMATION:
; APPLICANT: Bayer Aktiengesellschaft
; TITLE OF INVENTION: Ligand-gated anion channels of insects
; FILE REFERENCE: A 34 397
; CURRENT FILING DATE: 2001-03-14
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: DE 100 136 19.2
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 17
; LENGTH: 1257
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1254)
US-09-808-483-17

Query Match 1.3%; Score 19; DB 10; Length 1257;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 346 CATGCTGACCTGTGTGAC 364
|||||
Db 775 CATGCTGACCTGTGTGAC 793

RESULT 13
US-09-845-157-1/c
; Sequence 1, Application US/09845157
; Patent No. US20020090618A1
; GENERAL INFORMATION:
; APPLICANT: Smith, J.
; TITLE OF INVENTION: Thermostable Reverse Transcriptases and Uses Thereof
; FILE REFERENCE: 0942,5040001
; CURRENT FILING DATE: 2001-05-01
; PRIOR FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: US 60//207,196
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: M-MLV reverse transcriptase gene
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2151)
US-09-845-157-1

Query Match 1.3%; Score 19; DB 10; Length 2151;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ACCATGATGATGATGATGA 92
|||||
Db 33 ACCATGATGATGATGATGA 15

RESULT 14

```

US-09-965-703-74/C
: Sequence 74, Application US/09965703
: Patent No. US20020119521A1
: GENERAL INFORMATION:
: APPLICANT: Rohm and Haas Company
: APPLICANT: Paili, Subba Reddy
: APPLICANT: Kapitskaya, Marianna Zinovjevna
: APPLICANT: Cress, Dean Ervin
: TITLE OF INVENTION: No. US20020119521A1e1 Ecdysone Receptor-Based Inducible Gene Expr
: FILE REFERENCE: A01020B
: CURRENT APPLICATION NUMBER: US/09/965,703
: CURRENT FILING DATE: 2001-09-26
: PRIOR APPLICATION NUMBER: 60/191,355
: PRIOR FILING DATE: 2000-03-22
: PRIOR APPLICATION NUMBER: 60/269,799
: PRIOR FILING DATE: 2001-02-20
: PRIOR APPLICATION NUMBER: PCT/US01/09030
: PRIOR FILING DATE: 2001-03-21
: NUMBER OF SEQ ID NOS: 75
: SOFTWARE: Patentln version 3.1
: SEQ ID NO 74
: LENGTH: 3157
: TYPE: DNA
: ORGANISM: Escherichia coli
US-09-965-703-74

```

```

Query Match          1.3%; Score 19; DB 10; Length 3157;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 74 ACCGATGATGATGATGA 92
|||||
DB 33 ACCGATGATGATGATGA 15

```

```

RESULT 15
US-09-969-373-341/C
: Sequence 341, Application US/09969373
: Patent No. US2002013852A1
: GENERAL INFORMATION:
: APPLICANT: Eiertz, Roger J.
: APPLICANT: Hauge, Brian M.
: TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
: FILE REFERENCE: 38-10(52679)A
: CURRENT APPLICATION NUMBER: US/09/969,373
: CURRENT FILING DATE: 2001-10-02
: PRIOR APPLICATION NUMBER: US 09/754,853
: PRIOR FILING DATE: 2001-01-05
: PRIOR APPLICATION NUMBER: US 09/760,427
: PRIOR FILING DATE: 2001-01-13
: PRIOR APPLICATION NUMBER: US 09/855,768
: PRIOR FILING DATE: 2001-05-15
: NUMBER OF SEQ ID NOS: 4593
: SEQ ID NO 341
: LENGTH: 129
: TYPE: DNA
: ORGANISM: Glycine max
US-09-969-373-341

```

```

Query Match          1.2%; Score 18; DB 10; Length 129;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 76 CATGATGATGATGATGAA 93
|||||
DB 79 CATGATGATGATGATGAA 62

```

Search completed: November 29, 2002, 10:10:34
 Job time : 88 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2002, 10:09:13 ; Search time 2057 Seconds

(Without alignments)
11739.169 Million cell updates/sec

Title: US-09-868-760-6
Sequence: 1991
1 gavggggagacgtacga.....gcgtgtatagacgagca 1491

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size: 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

EST:
1: em_estba:*
2: em_esthm:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_estz:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vtl:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26	1.7	1101	17	CNS000HD3
2	25	1.7	529	14	BQ118156
3	23	1.5	541	9	AI724181
4	22	1.5	302	9	AU267147
5	22	1.5	303	9	AU269373
6	22	1.5	330	9	AI931386
					AI931386 ut69d11.y

7	22	1.5	500	17	AZ239345	AZ239345	RPCT-23-3
8	22	1.5	699	12	B1408775	B1408775	
9	21	1.4	171	12	BF759175	BF759175	RC4-CT047
10	21	1.4	368	13	B1702679	B1702679	IT56M03.Y
11	21	1.4	387	17	AQ845314	AQ845314	LMAUFV1-L
12	21	1.4	461	10	AW944020	AW944020	LP03320.3
13	21	1.4	465	13	BM159117	BM159117	1427412.Y
14	21	1.4	493	13	B1946460	B1946460	b5t59.5
15	21	1.4	649	10	AV610738	AV610738	AV610738
16	21	1.4	655	14	BP024062	BP024062	BP024062
17	21	1.4	691	17	AQ257939	AQ257939	nbx00119M
18	21	1.4	766	10	BE469502	BE469502	IPBdK0189
19	20	1.3	225	10	AW156710	AW156710	se29908.Y
20	20	1.3	244	12	BG311971	BG311971	UI-R-CVO-
21	20	1.3	292	17	AZ992197	AZ992197	2M0276E16
22	20	1.3	307	14	BQ980111	BQ980111	OGE10C16.
23	20	1.3	333	10	AV552114	AV552114	AV552114
24	20	1.3	344	12	BG511671	BG511671	sad09c05.
25	20	1.3	345	17	BH400144	BH400144	AC-ND-156
26	20	1.3	355	14	BQ986763	BQ986763	OGF10G05.
27	20	1.3	371	13	B1099490	B1099490	IP1.42.H0
28	20	1.3	375	14	B0539271	B0539271	MESt614-B
29	20	1.3	377	12	BF324534	BF324534	su25b06.Y
30	20	1.3	389	9	AU230857	AU230857	AU230857
31	20	1.3	404	10	BE523057	BE523057	MA23057M
32	20	1.3	410	12	BF521537	BF521537	EST458013
33	20	1.3	411	17	AZ86838	AZ86838	2M0177022
34	20	1.3	411	17	AQ402095	AQ402095	NS15068-A
35	20	1.3	412	17	BE592272	BE592272	WS15068-A
36	20	1.3	412	17	AZ619913	AZ619913	1M0452621
37	20	1.3	422	12	BG044255	BG044255	sa425h04.
38	20	1.3	443	10	BE125729	BE125729	DG1.54.G0
39	20	1.3	445	10	BE328931	BE328931	hg32d03.X
40	20	1.3	455	10	BE329345	BE329345	hg39g05.X
41	20	1.3	455	10	BE349696	BE349696	hg41g04.X
42	20	1.3	459	17	TA203C07Q	TA203C07Q	AL476545.T
43	20	1.3	460	14	BQ489299	BQ489299	73-E9180-
44	20	1.3	475	10	BB772499	BB772499	BB772499
45	20	1.3	480	13	BI273111	BI273111	NF092C06F

ALIGNMENTS

RESULT 1
CNS000HD3
LOCUS
DEFINITION
CNS000HD3
1101 bp
DNA
linear
GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence T63 end of BAC:
BACR34P05 of RPCT-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL073332.1 GI:4953112
GSS.
Drosophila melanogaster.
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Submitted (02-JUN-1999) Genoscope - Centre National de Sequenage:
BP 191 91006 EVRY cedex - FRANCE (E-mail: seqrefgenoscope.cns.fr
Web: www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazuhiro Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCT-98 and was constructed by partial

JOURNAL Unpublished (2002)
 COMMENT Contact: Hideko Urushihara
 Institute of Biological Sciences
 University of Tsukuba
 1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
 Tel: 81-298-53-4664
 Fax: 81-298-53-6614
 Email: hideko@bio1.tsukuba.ac.jp.
 Location/Qualifiers

FEATURES
 source 1. 302
 /organism="Dictyostelium discoideum"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="VSH255"
 /clone_1lb="VS"
 /sex="mat A"
 /dev_stage="vegetative"

BASE COUNT 127 a 39 c 53 g 82 t 1 others
 ORIGIN

Query Match 1.5%: Score 22; DB 9; Length 302;
 Best Local Similarity 100.0%; Pred. No. 3.6;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 50 ACCACGACGATGATCCACGATGA 71
 ||||||||||||||||||||
 Db 175 ACCACGACGATGATCCACGATGA 196

RESULT 5 303 bp mRNA linear EST 10-MAY-2002
 LOCUS AU269373
 DEFINITION AU269373 VS Dictyostelium discoideum cDNA clone VS1871 5', mRNA
 sequence.
 ACCESSION AU269373
 VERSION AU269373.1 GI:20528171
 KEYWORDS EST.
 SOURCE Dictyostelium discoideum.
 ORGANISM Dictyostelium discoideum.
 Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 1 (bases 1 to 303)
 Urushihara,H., Morio,T., Saito,T., Koriki,E., Ochiai,H., Maeda,M.,
 Takeuchi,I., Kohara,Y. and Tanaka,Y.
 Population analysis of cDNAs from unicellular and multicellular
 stages of Dictyostelium discoideum
 Unpublished (2002)
 Contact: Hideko Urushihara
 Institute of Biological Sciences
 University of Tsukuba
 1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
 Tel: 81-298-53-4664
 Fax: 81-298-53-6614
 Email: hideko@bio1.tsukuba.ac.jp.
 Location/Qualifiers

FEATURES
 source 1. 303
 /organism="Dictyostelium discoideum"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="VS1871"
 /clone_1lb="VS"
 /sex="mat A"
 /dev_stage="vegetative"

BASE COUNT 112 a 50 c 58 g 83 t
 ORIGIN

Query Match 1.5%: Score 22; DB 9; Length 303;
 Best Local Similarity 100.0%; Pred. No. 3.6;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 50 ACCACGACGATGATCCACGATGA 71
 ||||||||||||||||||||
 Db 194 ACCACGACGATGATCCACGATGA 215

RESULT 6 330 bp mRNA linear EST 30-JUL-1999
 LOCUS A1931386
 DEFINITION u169d11.y1 Sugano mouse kidney mKia Mus musculus cDNA clone
 IMAGE:2135829 5' similar to gb:J04806 Mus musculus osteopontin mRNA
 , complete cds (MOUSE);, mRNA sequence.
 ACCESSION A1931386
 VERSION A1931386.1 GI:5667940
 KEYWORDS EST.
 SOURCE Mus musculus.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 330)
 Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
 Underwood K., Stepien M., Pape, D., Harvey, N., Schurk, R., Ritter
 , E., Korn, S., Shih, T., Jackson, I., Cardenas, M., McCann, R.,
 Waterston, R. and Wilson, R.
 The WashU-NCI Mouse EST Project 1999
 Unpublished (1999)
 Contact: Maria M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LIND, contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:1000505
 Trace considered overall poor quality
 Seq primer: custom primer used
 High quality sequence stop: 1.
 Location/Qualifiers

FEATURES
 source 1. 330
 /organism="Mus musculus"
 /strain="C57BL"
 /db_xref="taxon:10090"
 /clone="IMAGE:2135829"
 /clone_1lb="Sugano mouse kidney mKia"
 /sex="Female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
 (CACGTGTC); Site_2: DraIII (CACCATGTC); 1st strand cDNA
 was primed with an oligo(dT) primer
 [ATGTGGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was
 ligated to a DraIII adaptor [TGTGGCCCTACTG], digested
 and cloned into distinct DraIII sites of the pME18S-FL3
 vector (5' site CACGTGTC, 3' site CACCATGTC). XhoI should
 be used to isolate the cDNA insert. Size selection was
 performed to exclude fragments <1.5kb. Library
 constructed by Dr. Sumio Sugano (University of Tokyo
 Institute of Medical Science). Custom primers for
 sequencing: 5' end primer CTCTGCTCTTAAGCTGCG and 3' end
 primer CGACTGCTGAGCTGACACA."

BASE COUNT 97 a 108 c 69 g 56 t
 ORIGIN

Query Match 1.5%: Score 22; DB 9; Length 330;
 Best Local Similarity 100.0%; Pred. No. 3.9;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 65 ACCATGACGACCATGATGATGA 86
 ||||||||||||||||||||
 Db 282 ACCATGACGACCATGATGATGA 303

RESULT 7 500 bp DNA linear GSS 15-JUN-2000
 LOCUS A239345
 DEFINITION RPCI-23-32B8.TJB RPCI-23 Mus musculus genomic clone RPCI-23-32B8.

DNA sequence.
 ACCESSION A2239345
 VERSION A2239345.1 GI:8552536
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 500)
 Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shatsman, S., Akintet, B., Levis, M., McGinn, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P., and Fraser, C. M.
 Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)
 Other_GSSs: RPCI-23-32B8.TVB
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@igf.org
 clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter.dejong@med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.ligr.org/cdb/bac_ends/mouse/bac_end_intro.html
 Plate: 32 row: B column: 8
 Seq primer: SP6
 Class: BAC ends.
 FEATURES
 source Location/Qualifiers
 1..500
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-32B8"
 /clone_1lb="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBAC63.6; Site: 1; EORI: Site.2; EORI: Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EORI and EORI Methylase. Size selected DNA was cloned into the pBAC63.6 vector at the EORI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."
 BASE COUNT 152 a 83 c 118 g 147 t
 ORIGIN
 Query Match 1.5%; Score 22; DB 17; Length 500;
 Best Local Similarity 100.0%; Pred. No. 4.7;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1292 ATGCTCATGATGCTCATGA 113
 ||||||||||||||||||
 Db 479 ATGCTCATGATGCTCATGA 500
 RESULT 8
 BJA08775/c 699 bp mRNA linear EST 10-MAR-2002
 LOCUS BJA08775 Dictyostellum discoidium cDNA library, SF Dictyostellum
 DEFINITION dictoidium cDNA clone dds46m2 3', mRNA sequence.
 ACCESSION BJA08775
 VERSION BJA08775.1 GI:19321559
 KEYWORDS EST
 SOURCE Dictyostellum discoidium.
 ORGANISM Dictyostellum discoidium.
 Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 1 (bases 1 to 699)
 Urushihara, H., Tanaka, Y., Kohara, Y., and Shin-I, T.
 Full length cDNA of Dictyostellum discoidium at the slug stage

JOURNAL Unpublished (2002)
 COMMENT Contact: Tadasi Shin-I
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.
 FEATURES
 source Location/Qualifiers
 1..699
 /organism="Dictyostellum discoidium"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="dds46m2"
 /clone_1lb="Dictyostellum discoidium cDNA library, SF"
 /sex="mat A"
 /dev_stage="Slug stage"
 BASE COUNT 188 a 93 c 154 g 262 t 2 others
 ORIGIN
 Query Match 1.5%; Score 22; DB 13; Length 699;
 Best Local Similarity 100.0%; Pred. No. 5.3;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 77 ATGATGATGATGATGAACAT 98
 ||||||||||||||||||
 Db 652 ATGATGATGATGATGAACAT 631
 RESULT 9
 BF759175/c 171 bp mRNA linear EST 12-JAN-2001
 LOCUS BF759175 RC4-CT0477-011200-016-e09 CT0477 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BF759175
 ACCESSION BF759175.1 GI:12107075
 VERSION BF759175.1
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 171)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., de Silva, W. J., Zago, M. A., Bordin, S., Costa, F. E., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H., Brunstein, A., deoliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J., and Simpson, A. J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 JOURNAL MEDLINE
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC4&t2=RC4-CT0477-011200-016-e09&t3=2000-12-01&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 4
 High quality sequence stop: 171.
 FEATURES
 source Location/Qualifiers
 1..171
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1lb="CT0477"
 /dev_stage="Adult"
 /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:

Mon Dec 2 12:19:25 2002

us-09-868-760-6.oli.rst

Page 5

SmaI. A mini-library was made by cloning products derived from ONESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 53 a 39 c 42 g 37 t

ORIGIN

Query Match 1.4%; Score 21; DB 12; Length 171;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 70 GAGCACCATGATGATGATGAT 90
|||||

Db 126 GAGCACCATGATGATGATGAT 106

RESULT 10
BI702679/c 368 bp mRNA linear EST 18-SEP-2001
LOCUS f156h03.y1 zebrafish SJD day 8 fin regeneration Danio rerio cDNA
DEFINITION Clone 4961885 5', mRNA sequence.
ACCESSION BI702679
VERSION BI702679.1 GI:15665308
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 368)
Clark, M., Johnson, S.L., Lebrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person, B., Waller, T., Gibbons, M., Pape, D., Harvey, N., Schuit, R., Rittler, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.
Washu zebrafish EST Project 1998
Unpublished (1998)
Other_ESTS: f156h03.x2
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbratfsh@wustl.edu
CDNA Library construction by: Joe Barnes and Steve Johnson, DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: Research Genetics web address: <http://www.researchgenetics.com/>
Putative full length read
The vector to vector length is 369
zebrafish identity (p-value greater than 1e-99) found to:
g1124464031gb1AA6057501AA605750 fa18fi2.s1 Ekerereally gastrulation zebrafish
Seq primer: 73 ET from AmerSham.

FEATURES
source
1..368
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="4961885"
/clone_lib="zebrafish SJD day 8 fin regeneration"
/sex="male"
/tissue_type="fin, 8-day regeneration"
/lab_host="DH10B"
/note="Vector: PAMPI. Site_1: EcoRI; Site_2: NotI; First strand cDNA synthesis was primed using oligo-dT on magnetic beads with an additional primer
5'-ggcgccgataatagcactacata-taggg-3'. Second strand synthesis was a 3-cycle PCR using the primers
5'-ggcgccgataatagcactacata-3' and
5'-aagcgctgtgtacacacgcagagactctt-tttttttttt-3'. cDNA

was subsequently amplified in a 7-cycle PCR with the following primers: 5'-ggcgccgataatagcactacata-3' and 5'-aagcgctgtgt-tacacacgcag. Deoxy-OMP adaptors were added in a third PCR (5 cycles) and the primers
5'-caucaucaucaagcgccgataatagcactacata-3' and
5'-cucacucaucaagcgctgtgtacacacgcag-3'. Ends were treated with uracil DNA glycosylase and product with 3' overhangs was annealed to complementary ends of PAMPI. Insert can be excised using EcoRI and NotI. Library constructed by Joe Barnes and Steve Johnson (Washington University)."

BASE COUNT 122 a 67 c 80 g 99 t

ORIGIN

Query Match 1.4%; Score 21; DB 13; Length 368;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 72 CGACCATGATGATGATGATGATCA 92
|||||

Db 32 CGACCATGATGATGATGATGATCA 12

RESULT 11
AO845514/c 387 bp DNA linear GSS 25-MAY-2001
LOCUS LMAJFV1_lm26e03.y1 Leishmania major FV1 random genomic library
DEFINITION Leishmania major genomic clone LMAJFV1_lm26e03 5', similar to contigs element VI-chl_type_II.2 leishmania repetitive element ;, DNA sequence.
ACCESSION AO845514
VERSION AO845514.1 GI:6050162
KEYWORDS GSS.
SOURCE Leishmania major.
ORGANISM Leishmania major.
Eukaryota; Euzenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
1 (bases 1 to 387)
Akopyants, N.S., Clifton, S.W., Martin, J., Pape, D., Wylie, T., Li, L., Kissinger, J., Roos, D.S., Marra, M., Hillier, L., Chinwalla, A., Blaisdell, A., Schmitt, A., Person, B., Theising, B., Rittler, E., Ronko, I., Bennett, J., Cole, R., Underwood, K., Cardenas, M., Gibbons, M., Harvey, N., McCann, R., Tsagaris, V., Williams, T., Jackson, Y., Bowers, Y., Waller, T., Waterston, R., Wilson, R., and Beverley, S.M.
A survey of the Leishmania major Friedlin strain VI genome by shotgun sequencing: a resource for DNA microarrays and expression profiling
Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)
21192569
Other_GSSs: lm26e03.x1
Contact: Akopyants, NS / Beverley, SM
Washu Leishmania Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Library construction: Natalia S. Akopyants, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
If using this information please cite:
N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major Friedlin strain VI genome by shotgun sequencing' and the Washington University Genome Sequencing Center for information on obtaining clone material please contact: Natalia S. Akopyants Ph.D.
(natalia@borcim.wustl.edu) and/or Stephen M. Beverley Ph.D.
(beverley@borcim.wustl.edu)
Seq primer: -40RP from gibco
Class: shotgun
High quality sequence stop: 370.

FEATURES
source
1..387
/organism="Leishmania major"
/strain="Friedlin strain VI"

```

/db.xref="taxon:5664"
/clone_1lib="1m26s03"
/lab_host="10p10 (Invitrogen)"
/notes="Vector: pZero-2 (Invitrogen); Site_1: EcoRV;
Genomic DNA was isolated from stationary phase cells. For
this library, DNA was sheared to give a tight size
distribution of 1-1.5kb fragments, blunt-ended with T4 DNA
polymerase, dephosphorylated with Shrimp Alkaline
Phosphatase and ligated into pZero-2 vector's EcoRV site."
BASE COUNT      42 a      121 c      93 g      129 t
ORIGIN
Query Match      1.4%; Score 21; DB 17; Length 387;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      31 AAGATGACACCATGACGAC 51
|||||
Db      299 AAGATGACACCATGACGAC 279

RESULT 12
LOCUS      AM944020/c      461 bp      mRNA      linear      EST 23-APR-2001
DEFINITION      LP03320.3prime LP Drosophila melanogaster larval-early pupal p072
                  Drosophila melanogaster cDNA clone LP03320 3 similar to AC004433;
                  FBP0002727 sk1 FBN0016984 FLYBASE:FBP0002727
                  AC004433:6748..9126:03/15/2001, mRNA sequence.
ACCESSION      AM944020
VERSION      AM944020.1 GI:8121748
KEYWORDS      EST.
SOURCE      fruit fly.
ORGANISM      Drosophila melanogaster
                  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
                  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
                  Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 461)
AUTHORS      Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
                  Lewis,S. and Rubin,G.M.
JOURNAL      BDGP/HMT Drosophila EST Project
COMMENT      Unpublished (2001)
                  Other ESTs: LP03320.3prime
                  Contact: Stapleton, M.
                  BDGP
                  Lawrence Berkeley National Lab
                  One Cyclotron Rd, Berkeley, CA 94720, USA
                  Fax: 510 486 6798
                  Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
                  Based upon the presence of a XhoI site followed by a run of 14 or
                  more T residues at the beginning of the sequence, this clone was
                  polyadenylated. The resulting Poly-T sequence has been removed. hit
                  genomic AE003452: Drosophila melanogaster genomic scaffold
                  14200013386038 section 1 of 15, complete sequence.: 03/15/2001
                  Plate: LP.33 row: B column: 8
                  High quality sequence stop: 367.
FEATURES
source
1..461
Location/Qualifiers
/organism="Drosophila melanogaster"
/db.xref="taxon:7227"
/clone_1lib="LP03320"
/clone_1lib="LP Drosophila melanogaster larval-early pupal
p072"
/sex="male and female"
/dev_stage="larvae-pupae"
/lab_host="DH5-alpha"
/notes="Organ: whole body; Vector: p072; Site_1: EcoRI;
Site_2: XhoI; Sized fractionated cDNAs were directly
ligated into p072. Plasmid cDNA library."
BASE COUNT      82 a      132 c      119 g      128 t
ORIGIN

```

```

Query Match      1.4%; Score 21; DB 10; Length 461;
Best Local Similarity 100.0%; Pred No. 15;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      37 GACCACCATGACGACCCAC 57
|||||
Db      281 GACCACCATGACGACCCAC 261

RESULT 13
LOCUS      BM159117/c      465 bp      mRNA      linear      EST 03-DEC-2001
DEFINITION      f27d12.y1 zebrafish C32 14 somite embryo Danio rerio cDNA clone
                  5565767 5' similar to SW:YZAL_HUMAN Q16465 HYPOTHETICAL PROTEIN ;
                  mRNA sequence.
ACCESSION      BM159117.1 GI:17247883
VERSION      BM159117
KEYWORDS      EST.
SOURCE      zebrafish.
ORGANISM      Danio rerio
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
                  ; Cyprinidae; Danio.
REFERENCE      1 (bases 1 to 465)
AUTHORS      Clark,M., Johnson,S.L., Lehnach,H., Lee,R., Li,F., Marra,M., Eddy
                  S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
                  K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
                  Waller,T., Gibbons,M., Page,D., Harvey,N., Schurk,R., Ritter,E.,
                  Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
                  and Wilson,R.
JOURNAL      WashU zebrafish EST Project 1998
COMMENT      Unpublished (1998)
                  Contact: Stephen L. Johnson
                  Washington University School of Medicine
                  444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
                  Tel: 314 286 1800
                  Fax: 314 286 1810
                  Email: zbrfish@wustl.edu
                  cDNA library construction by: Joe Barnes and Steve Johnson. DNA
                  sequencing by: Washington University Genome Sequencing Center Clone
                  distribution: Research Genetics web address:
                  http://www.researchgenetics.com/
                  Possible reversed clone: similarity on wrong strand
                  zebrafish identity (P-value greater than 1e-99) found to:
                  g112549301gb1AA6587761AA6587761a66a10.s1 zebrafish fin day3
                  regeneration
                  Seq primer: T3 ET from Amerisham.
FEATURES
source
1..465
Location/Qualifiers
/organism="Danio rerio"
/strain="C32"
/db.xref="taxon:7955"
/clone_1lib="5565767"
/clone_1lib="zebrafish C32 14 somite embryo"
/tissue_type="embryo, 14 somite"
/notes="Vector: pAMP1; Site_1: EcoRI; Site_2: NotI; First
strand cDNA synthesis was primed using oligo-dT on
magnetic beads with an additional primer
5'-ggcgccgtatagctatagc-tatagg-3'. Second strand
synthesis was a 3 cycle PCR using the primers
5'-ggcgccgtatagctatagc-tatagg-3' and
5'-aagcagtggtggt-aacaacgagcagc-3'. cDNA
was subsequently amplified in a 7-cycle PCR with the
following primers: 5'-ggcgccgtatagctatagc-3'
5'-aagcagtggtggt-aacaacgagcagc-3' and the primers
5'-cauacuaucuaaagcagcgtatagcgtatagc-3' and
5'-cauacuaucuaaagcagcgtatagcgtatagc-3'. Ends were
treated with uracil DNA glycosylase and product with 3'
overhangs was annealed to complementary ends of pAMP1.
Insert can be excised using EcoRI and NotI. Library
constructed by Joe Barnes and Steve Johnson (Washington

```

```

BASE COUNT      128 a      97 c      104 g      123 t      13 others
ORIGIN
Query Match      1.4%; Score 21; DB 13; Length 465;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 CGACCATGATGATGATGATGA 92
    |||||||
Db 22 CGACCATGATGATGATGATGA 2

RESULT 14
BI946460      493 bp      mRNA      linear      EST 19-OCT-2001
LOCUS      BEST59.5' Honeybee brain cDNA library Apis mellifera cDNA 5', mRNA
DEFINITION
ACCESSION      BI946460
VERSION      BI946460.1 GI:16285147
KEYWORDS      EST.
SOURCE      honeybee.
ORGANISM      Apis mellifera
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
Apoidea; Apis.
REFERENCE      1 (bases 1 to 493)
AUTHORS      Kucharski, R. and Maleszka, R.
TITLE      Evaluation of differential gene expression during behavioral
JOURNAL      development in the honeybee using microarrays and northern blots
MEDLINE      Genome Biol. 3 (2), research007.1-9 (2002)
COMMENT      21853287
Contact: Maleszka R
Visual Sciences Group
Research School of Biological Sciences, Australian National
University
Biology Pl., Canberra ACT 0200, Australia
Tel: +61 2 6125 0451
Fax: +61 2 6125 3784
Email: maleszk@rsbs.anu.edu.au
Seq primer: M13 reverse.
Location/Qualifiers
1. 493
/organism="Apis mellifera"
/db_xref="taxon:7460"
/clone_id="Honeybee brain cDNA library"
/sex="female"
/tissue_type="central brain, mushroom bodies, central body
, deutocerebrum"
/dev_stage="adult worker"
/lab_host="E.coli SOLR"
/note="Organ: brain; Vector: pBluescript II SK-; Site_1:
EcoRI; Site_2: XhoI; cDNA-harboring phagemids were
obtained by performing mass excision procedure on honeybee
brain cDNA lambda UniZap XR library (Ebert P.R. et al.,
1998, Insect Mol. Biol. 7(2):151-62)."
```

```

ACCESSION      AV610738
VERSION      AV610738.1 GI:9746408
KEYWORDS      EST.
SOURCE      Bos taurus
ORGANISM      Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE      1 (bases 1 to 649)
AUTHORS      Takasuga, A., Hirotsune, S., Itoh, R., Jitohzono, A., Suzuki, H., Aso, H.
and Sugimoto, Y.
TITLE      Establishment of a high throughput EST sequencing system using
JOURNAL      poly(A) tail-removed cDNA libraries and determination of 36,000
MEDLINE      bovine ESTs
COMMENT      Nucleic Acids Res. 29 (22), E108 (2001)
21570554
Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Ogikura, Nishigo, Nishi-Shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@ococa.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a poly(A)-deleted cDNA library.
Location/Qualifiers
1. 649
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_id="EILU034C11"
/clone_id="Bos taurus lung fetus"
/tissue_type="lung"
/dev_stage="fetus"
/lab_host="DH10B"
/note="Vector: pZL1; Site_1: SalI; Site_2: NotI; Poly A
was deleted from a NotI site"
BASE COUNT      185 a      118 c      187 g      159 t
ORIGIN
Query Match      1.4%; Score 21; DB 10; Length 649;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 GACCATGATGATGATGATGA 93
    |||||||
Db 591 GACCATGATGATGATGATGA 611
```

Search completed: November 29, 2002, 12:31:13
Job time : 2078 secs

```

RESULT 15
AV610738      649 bp      mRNA      linear      EST 28-NOV-2001
LOCUS      AV610738 Bos taurus lung fetus Bos taurus cDNA clone EILU034C11 5',
DEFINITION      mRNA sequence.
```

THIS PAGE BLANK (USPTO)